The theoretical and practical value of the OECD policy advice for education

ALOISI, CESARE

The theoretical and practical value of the OECD policy advice for education

Cesare Aloisi

Abstract

This research seeks to evaluate the policy advice on educational matters provided by the Organisation for Economic Co-operation and Development (OECD), particularly through its Programme for International Student Assessment (PISA). The thesis assesses the "value" of the OECD education policy advice (EPA), that is, whether the EPA is advice worth following because there is evidence that it *could help* or *has helped* countries to achieve the goals of greater quality and equity in education.

Firstly, a multi-agent practical reasoning model and political discourse analysis are applied to OECD publications to extract education policy recommendations. The EPA is thus comprehensively reconstructed, summarised and organised. The EPA is analysed in terms of its internal coherence, its consistency over time and over publications, as well as its soundness: whether there is empirical evidence that what the OECD suggests that countries should do could work in principle.

In the second part of the thesis, evidence of effective EPA-aligned country policies is sought using national and cross-national level data and a mixed-methods approach. From a cross-national perspective, country alignment with the EPA is analysed by collecting 61 expert judgements from 30 PISA countries using a specifically-developed online questionnaire. The plausible impact of EPA-aligned policies is inferred by applying multilevel growth models to PISA trends in performance and equity.

The cross-country analysis is complemented by two case studies: The Republic of Ireland and the French Community of Belgium. Data on 50 policies is collected from more than 800 policy documents and through 25 semi-structured expert interviews carried out over two study visits. Policy alignment is analysed with a visual adaptation of the delta-convergence method to account for qualitative data. The effectiveness of EPA-aligned policies is evaluated using a framework developed by Prof J. Douglas Willms and adopted by the OECD, as well as national evaluation reports.

The theoretical and practical value of the OECD policy advice for education

What is the evidence that integrating OECD recommendations into national policies could help and has helped countries to achieve the goals of greater quality and equity in education?

Volume 1 of 2

Cesare Aloisi

A thesis submitted in partial fulfilment of the qualification of

Doctorate of Philosophy in Education

School of Education, Durham University

2016

Table of Contents

List of tables

List of figures

List of abbreviations and formatting conventions

List of the main abbreviations:

Abbreviation	Definition
ACER	Australian Council for Educational Research
AfL	Assessment for learning
ASPE	service for the Analysis of Systems and Practices in Education
BERA	British Educational Research Association
CEB	*Certificat d'Études de Base*, basic learning certificate
CEEE	Centralised external exit examination
CEF	*Conseil de l'éducation et de la formation*
CERI	Centre for Educational Research and Innovation
CFB	*Communauté française de Belgique*, French Community of Belgium
CPD	Continuing professional development
CSQ	Case-study question
DAS	Disadvantaged Areas Scheme
DEIS	Delivering Equality of Opportunity in Schools
DES	Department of Education and Skills
EAG	"Education at a Glance"
EENEE	European Expert Network on Economics of Education
EPA	Education policy advice
ERC	Educational Research Centre
ERT	European Round Table of Industrialists
ETS	Education Testing Service
Forfás	National Policy and Advisory Board for Enterprise, Trade, Science, Technology and Innovation
GCEB	Giving Children an Even Break
GDP	Gross domestic product
GER	Gross enrolment ratio
HSCL	Home/School/Community Liaison scheme
IAEP2	Second International Assessment of Educational Progress
IALS	International Adult Literacy Survey
IEA	International Association for the Evaluation of Educational Achievement
INTO	Irish National Teachers' Organisation
IO	International / intergovernmental organisation
ITT	Initial teacher training
JCE	Junior Certificate Examination
lNES	Indicators of Education Systems
LSA	Large-scale assessment
MoE	Ministry/Minister of Education
NAMER	National assessments of reading and mathematics
NCCA	National Council for Curriculum and Assessment
NLNS	National Literacy and Numeracy Strategy
NPM	National Project Manager

OECD	Organisation for Economic Co-Operation and Development
OEEC	Organisation for European Economic Co-operation
PAIQ	Policy Advice Implementation Questionnaire
PGB	PISA Governing Board
PIRLS	Progress in International Reading Literacy Study
PISA	Programme for International Student Assessment
PO	*Pouvoirs organisateurs*, organising powers
rJCMS	revised Junior Certificate Mathematics Syllabus
rJCSS	revised Junior Certificate Science Syllabus
RLS	Reading Literacy Study
RQ	Research question
SEN	Special education needs
SES	Socio-economic status
SSE	School self-evaluation
STEM	Science, technology, engineering and mathematics
TALIS	Teaching And Learning International Survey
TIMSS	Trends in International Mathematics and Science Study
UNESCO-IBE	United Nations Educational, Scientific and Cultural Organization - International Bureau of Education
UNESCO-UIS	United Nations Educational, Scientific and Cultural Organization - UNESCO Institute for Statistics
WSE	Whole school evaluations

Formatting and language conventions:

Formatting	**Use**
'inverted commas'	A quotation
"double inverted commas"	A concept (e.g., "Education policy advice" is…) A word used in a non-standard sense (e.g., PISA has become a well-known "brand" among policymakers) A term embraced by others but not by this author (e.g., "neo-liberal" education policies)
« *guillemets* »	They are the French equivalent of the double inverted commas; in this document they are used to name legislation, following the convention in the CFB (e.g., *Décret « Missions »)*
Italic	A publication title A word not in English The name of a statistical variable A way to emphasise a word

Declaration

The material contained in the thesis was never submitted for a degree in this or any other institution; the parts of the thesis on pre-primary education in Chapter 3 and Chapter 4 were expanded and adapted into a research article which is currently under review at the *Large-Scale Assessments in Education* journal.

Statement of copyright

Acknowledgements

I wish to record my gratitude to my supervisors Peter Tymms, Christine Merrell (second and third year), and Judith Glaesser (first year). Their feedback was always comprehensive and insightful, and their guidance has been professional and considerate.

This thesis would not have been possible without the hospitality of Peter Archer at the Educational Research Centre in Dublin, and Dominique Lafontaine at the *Service d'Analyse des Systèmes et des Pratiques d'Enseignement* in Liège. I would like to thank both research teams for their time and their friendliness, as well as all the experts who provided data for this work through interviews and questionnaires.

I am grateful to all the staff at Ustinov College for its kindness and approachability. Special words of appreciation are due to Rebecca Bouveng and Theresa McKinven for involving me in the Global Citizenship Programme, which was an opportunity to develop my skills and make new friends, and provided key financial support.

Similarly, I would like to extend my thanks to the team led by Rob Coe at the Centre for Evaluation and Monitoring at Durham, for the many research collaborations which introduced me to the beautiful Rasch model for educational measurement; and to Steve Higgins, for the fruitful collaboration and the pleasant conversations.

My stay in Durham was made ever more enjoyable by some great friends I had the pleasure to get to know. Here, I would like to name amongst others Erika Taretto, Chiara Grigolin, Rumy Grozdanova, Henry Jones and Jane Rooney; and in particular, Ntina Tzouvala and Massimilano Rota, for helping me to solve social issues.

I would also like to acknowledge the role of Alice Panepinto, whose presence always seem to make my day better than it would otherwise be (though surely this might just be a correlational effect!).

Finally, I am forever indebted (quite literally) to my family and especially my mother Ombretta Campioli, who has given me unconditional love and support in all these years. Her work and the work of my father Giuseppe Aloisi have substantively improved my life trajectory. A huge good luck to my brother Nicolò Aloisi for his professional career.

Cesare Aloisi

This thesis was proofread by Matt Hann. The HTML code for the OECD Policy Advice Implementation Questionnaire was reviewed and improved by Riccardo Galli. The multilevel model diagnostics were reviewed by Adetayo Kasim.

Dedication

To my nonna Silvana, the only family member who actually bothered to look up Durham on the map.

Volume 1: Main analysis

Volume 1 of 2

Introduction

This research seeks to evaluate the policy advice on educational matters (education policy advice, EPA) provided by the Organisation for Economic Co-operation and Development (OECD). It does this by assessing the extent to which following OECD recommendations can help countries to achieve the goals of greater quality and equity in education. The evaluation analyses the theory and practice of the OECD policy advice: on the one hand, what the OECD says countries should do and whether this is reasonable advice; on the other hand, which recommendations find actual application on a national level and whether this can be linked to improvements in quality (i.e., student performance) and equity.

The rationale for the evaluation is that, in recent years, the OECD has gained prominence in national educational matters, particularly after the introduction of the Programme for International Student Assessment (PISA) in 2000. PISA has become a well-known "brand" among policymakers, and as a result, participating countries have started to introduce policies in response to their performance in the assessment. This has led to a strand of research investigating the use, usefulness and effects of PISA: how different countries prepare and respond to the survey; whether it is a valid and reliable measure of student abilities; and how international surveys such as PISA promote a culture of accountability at the country level.

By placing the spotlight on PISA, this literature has generally left the OECD policy advice unanalysed, or it has identified and critiqued only its most distinctive "neo-liberal" aspects (deregulation and decentralisation). Many national policies claiming to be in line with OECD recommendations have been taken at face value as examples of the globalising effects of neo-liberalism on education systems through PISA. However, a systematic and comprehensive analysis of what the OECD actually advises that countries should do to improve education is currently missing. Therefore, it is unclear whether policies purporting to be aligned with the EPA are truly so, and whether this has helped countries to increase the performance and equity of their education systems as envisaged by the OECD.

This research is an attempt to fill these gaps by moving the focus of the investigation from PISA to the EPA by assessing the quality of the EPA and the ability of EPA-informed policies to raise country performance in the three PISA literacies of Reading, Mathematics and Science[1] and country equity in education. If the EPA is advice worth following, it should be possible to detect improvements at a national or international level when it is followed. In this introductory section, a brief outline of each chapter is provided as a guide to the reader.

[1] The capitalisation is used in this document to differentiate OECD-defined constructs from other definitions of "reading", "mathematics" and "science", see Chapter 1.

Chapter 1 opens with a historical account on the evolution of the OECD aimed at showing its growing influence in national educational decisions among and beyond its member states. Once relegated to a subordinate role, education has become a key OECD activity since the Indicators of Education Systems (INES) programme in 1988 and PISA ten years later (Dale & Robertson, 2007), which allowed the OECD to produce education data autonomously. The chapter presents the complicated structure behind the development, administration and strategy of PISA, paying particular attention to the role of the Secretariat (specifically the Directorate for Education and Skills).

The Directorate is both a network of researchers and the administrative arm of an organisation at the service of member countries. It has the autonomy to carry out policy-relevant research, and sufficient control over information to be in a powerful position when it comes to providing education policy advice. However, it lacks absolute ownership of the knowledge it produces, and it is characterised by a persistent tension between the creative, scientific and epistemic work expected from the highly-qualified staff employed at the organisation, and the reality of having to satisfy 'externally defined national interests for utility maximization' (Marcussen & Trondal, 2011, p. 611).

Central to Chapter 1 is the argument that PISA is not simply a diagnostic tool, a monitoring instrument to measure the state of education systems. Alongside a more "descriptive" (though not necessarily neutral) side, comprising country statistics and outcomes from the assessment, PISA has a strong normative component—the EPA—telling participants what policy options are recommended given the evidence from the survey.

PISA might be thought of as a "public policy instrument", 'a device that is both technical and social, […] carrying a concrete concept of the politics/society relationship and sustained by a concept of regulation' (Lascoumes & Le Galès, 2007, p. 4). It is argued that while the quality of the technical-descriptive side of PISA is linked to the OECD's ability as a test developer, the quality of the socio-normative aspect is linked to its credibility as a policy advisor. This is what makes the EPA a critical (but under-researched) component of the OECD activity in education.

The chapter continues with a review of PISA / EPA-informed policymaking on a national level. This includes country motivation to be assessed by PISA and the reception and uses of the survey by its participants, from the "prodigal child" Finland to the (formerly) "problem child" Germany. The evidence suggests that the technical-descriptive side of PISA has both a use and an impact within countries, whereas it is not clear whether this is the case for its socio-normative side—i.e., whether countries take notice of and enact OECD recommendations.

Chapter 1 concludes with an explanation of the purpose, timeliness and relevance of this research. Moving the focus from PISA to the EPA means addressing a gap in the literature

and looking at the issue from a broader perspective, one that sees PISA as one component within a larger OECD strategy to produce, disseminate and replicate educational knowledge.

Chapter 2 introduces the theoretical framework and the mixed-methods design of this research, as well as its leading questions and guiding hypotheses. The study can be framed as an exercise in education policy advice evaluation, aiming to contribute to the fields of policy evaluation and (international) assessment validation. From validation theory, the research borrows the concepts of "validity" and "reliability", and it applies them to PISA; from evaluation theory, the research borrows the concept of "programme theory" (Chen, 1990), and applies it to the EPA. It also agrees with realistic evaluation (Pawson & Tilley, 1997) that no policy 'just works': policy evaluation has to take the context of implementation into account when looking for policy effects. Chapter 2 expands on the contribution of both fields and clarifies how they are linked by and within this document.

Because of the central role of the EPA, a working definition of "education policy advice" is given in the chapter. The form it might take in OECD publications (in terms of rhetoric style and textual features) is reconstructed drawing from practical argumentation theory (Fairclough & Fairclough, 2012; Walton, 2006a). The EPA is considered the endpoint of an argument that starts with a tension between the goals and values of the OECD and some problematic factors which prevent their achievement. The EPA is the proposed solution for the clash, and PISA is just one of the pieces of evidence supporting the argument.

The research is an attempt to relate a complex knowledge-production process (PISA-informed EPA), via its chain of implications at the national level (the development and implementation of EPA-informed policies), to the endpoint: the effectiveness of national policies in raising student performance and educational equity. Each of these three aspects is associated with a research question, and a fourth one guides the conclusions to the thesis.

To answer the first research question (RQ1), the EPA is reconstructed in Chapter 3 in order to understand what the OECD says countries should do and to ask whether this is reasonable advice. Firstly, a policy advice "blueprint" developed from practical argumentation theory is applied to selected OECD publications in an iterative fashion to reconstruct the EPA. OECD recommendations are identified and thematically organised. The organisation of the OECD policy advice in one comprehensive strategy represented by

Table 3.2 integrates the feedback of the Director of PISA, Andreas Schleicher, and it provides the bases for evaluating country policy alignment and effectiveness in the following chapters.

In the continuation of Chapter 3, the EPA is validated along three dimensions: internal coherence, consistency over time and soundness. In the context of this research, "soundness" is not about truthfulness or falsehood, but rather about the extent to which the EPA is supported by evidence. In other words, what has the OECD been basing its policy advice on? And how does the OECD know that what it suggests will produce any positive effects?

Since it would have been unfeasible to evaluate every single policy recommendation, a few were selected from the four main policy areas in the EPA (educational time, equity, quality of provision and performance). Each of the policy levers selected for review represents a specific issue with the way the OECD grounds and delivers advice. The analysis covers pre-primary education (educational time); gender, streaming and school choice (equity); teacher professional development and school accountability (quality of provision) and curricular changes (performance). Problematic aspects are summarised and general conclusions are drawn at the end of the chapter.

In Chapter 4, evidence of country alignment with OECD recommendations (RQ2) and of the effectiveness of EPA-aligned policies (RQ3) is sought from a cross-country perspective. A cross-national measure of country alignment with PISA recommendations (the EPA Index) is introduced, and different approaches are used to estimate, albeit indirectly, the effectiveness of following the EPA. As part of the evaluation, the validity and the limitations of PISA as an instrument to measure change are also addressed.

The chapter employs both original data analyses (correlations, individual country regressions and multilevel models) and findings from other studies. It proceeds in three steps, which use increasingly targeted data. The first step models country outcomes of performance and equity over time using PISA data; the second relates policy outputs (i.e., measurable policy effects) to PISA outcomes; and the third links policy alignment with these outcomes.

The first step analyses trends in performance and equity since PISA started in 2000, with an expectation that if the EPA is adopted, implemented and effective, then some increases in performance or equity should be observed. The second step attempts to set an empirical limit on the extent to which education reforms can be expected to produce observable changes in country outcomes. This is done by contrasting the "effects" of three policy outputs (general economic productivity, equity levels and access to pre-primary education) on PISA scores. All three outputs are known to correlate to PISA outcomes, but the hypothesis is that only the indicator capturing a specific educational intervention (access to pre-primary education)

would be related to changes in PISA scores over time. The magnitude of these changes is the empirical ceiling of policy influences on PISA mentioned above.

The third step complements secondary PISA data analyses with information collected through expert questionnaires, to relate country levels of (and changes in) achievement and equity to their degree of alignment with OECD recommendations.

Finally, the chapter considers validity and reliability issues in relation to PISA, in particular issues affecting the comparability of scores between countries and over time. While the PGB decided that 'the establishment of reliable trends should become the overriding priority' (Mazzeo & von Davier, 2008, p. 4), PISA is still affected by 'a plethora of inaccuracies' (Wuttke, 2007, p. 261) that compromise by an unknown degree the accuracy of the reported scores and the validity of the statistical procedures. The OECD response to some of the PISA shortcomings is discussed, and conclusions are drawn.

Some authors (Baird et al., 2011; Pedró, 2012) have noted that the policy effects of PISA—and the value of the EPA—can only be appreciated through trend analyses and triangulation of different evidence sources. Therefore, in this thesis the cross-national evidence from Chapter 4 is complemented by evidence from two case studies (the Republic of Ireland and the French Community of Belgium, CFB) in Chapters 5 and 6.

In Chapters 5 and 6, which have the same structure, the use (RQ2) and impact (RQ3) of OECD education policy advice in Ireland and the CFB are analysed. The chapters are guided by case-study-specific questions and make use of policy documents and semi-structured expert interviews to gather data. In the first part, a historical overview of the main educational issues before the first PISA cycle frames the context in which the assessment was introduced. Irish and Belgian relations with the OECD, participation in other large-scale assessments (LSA) and the key educational legislative frameworks are also analysed.

At the national level, PISA is administered by national project managers, the individuals or institutions that who carry out the test, analyse the data and produce national reports. In Ireland and the CFB the national project managers are research centres, which are arguably in a similar position to that of the Directorate for Education and Skills at the OECD: their relevance comes not only from the quality of the research they produce, but also from their ability to inform policy. Therefore, it is hypothesised that their interpretation of PISA outcomes and consequent policy advice might compete with that of the OECD. To account for this possibility, Chapters 5 and 6 include a comparison of the policy recommendations for Ireland and the CFB from both the international version of the PISA report, authored by the Directorate for Education and Skills at the OECD, and the national one, authored by the national project managers.

The chapters proceed with a detailed discussion of selected policies covering the context surrounding the interventions, their degree of alignment with the EPA and, when possible, any evidence of implementation and impact. Three areas of reform are covered: curriculum, evaluation and assessment, and equity. Curricular reforms have the potential to impact on PISA performance directly, and reforms affecting equity outcomes are also analysed. Interventions relating to systemic evaluation and assessment are included as they represent the typical accountability policies that some research strands claim to be driven by globalising forces such as PISA. The case studies provide a good opportunity to understand whether these policies had been borrowed, what role the local context had in shaping them and how they could be linked to PISA.

Chapters 5 and 6 conclude with a summary of trends in policy convergence towards the OECD goals, and an evaluation of the plausible impact of national reforms on country outcomes in PISA. Policy alignment with the EPA is summarised, drawing from the concept of "delta-convergence", that is, 'the decreasing distance of policies towards an exemplary model, for example a model promoted by an international organisation' (Holzinger, Knill, & Arts, 2008, p. 83; see also Jakobi & Teltemann, 2009).

The delta-convergence value is a useful summary measure of the degree of EPA-alignment, and it has already found application in research on PISA (Bieber & Martens, 2011), but the 'exemplary model' provided by the EPA does not suit a quantitative indicator. Therefore, only the principles of delta-convergence are retained to produce a visual representation of the progressive alignment of national policies to the EPA.

Finally, in Chapters 5 and 6, the impact of national reforms on PISA outcomes is analysed using Willms' (2006) LSA-based policy evaluation framework. The framework divides education policies into categories: 'universal' interventions seek to raise the achievement of all students; 'SES-targeted' policies intervene on the performance of students with a low socio-economic status (SES); whereas 'performance-targeted' interventions try to improve the results of low achievers. Willms' framework is often used by the OECD as a means to evaluate policy effectiveness. The variations of indicators that are specific to each category are claimed to be related to specific policies. Here, Willms' framework is employed but not taken as an infallible instrument, and its ability to detect policy effects on PISA is also critiqued as part of the analysis.

Following the six main chapters, some limitations of this study are acknowledged and suggestions for future directions are given. A fourth Research Question (RQ4) helps to summarise findings from all previous analyses under one heading: how valuable is the OECD education policy advice? The answer to this question concludes the thesis.

Chapter 1: A literature review on the OECD, PISA and their influence on national policymaking

This chapter is articulated in the following parts:

- A brief history of the OECD and the increasingly prominent role of the OECD in education.
- A presentation of PISA and how it is governed and administered.
- The descriptive and socio-normative functions of PISA.
- The use of PISA and of OECD policy recommendations in participating countries
- The purpose and relevance of this research.

It is argued that countries use PISA rankings and indicators selectively to back policy positions; however, PISA also aims to be a public policy instrument, making recommendations to influence country policies. The literature has mostly focused on the first aspect of PISA, whereas this research will concentrate on the value of PISA-based education policy advice (EPA). The EPA is deemed a key but under-researched side of PISA because it is related to the credibility of the OECD as a policy advisor.

What is the OECD?

The Organisation for Economic Co-Operation and Development (OECD) was established on 30 September 1961 after a convention was jointly signed on 14 December 1960 by the Organisation for European Economic Co-operation (OEEC), the United States and Canada ("Convention on the Organisation for Economic Co-operation and Development," 1960). Born in 1948 to implement the Marshall Plan in Europe, the OEEC struggled to make members' positions converge, especially with respect to economic and trade agreements. When fears over Soviet technological advances accelerated following the launch of the Sputnik in 1957, the reshaping of the OEEC into the OECD was seen as an economic and political alternative to the expansion of NATO as a means to tighten Western links over the Atlantic (Carroll & Kellow, 2011).

The OECD is an international and intergovernmental organisation (IO) with legal personality[2] concerned with the promotion of policies 'that will improve the economic and

[2] The OECD inherited the legal personality possessed by the OEEC ("Convention on the Organisation for Economic Co-operation and Development," 1960), which was as follows: 'The Organisation shall possess juridical personality. It shall have the capacity to conclude contracts, to acquire and dispose of movable and immovable property and to institute legal proceedings.' ("Supplementary Protocol No. 1 to the Convention for European Economic Co-operation on the Legal Capacity, Privileges and Immunities of the Organisation," 1948, Art. 1).

social well-being of people around the world' (OECD, 2011a, p. 8). Topics of interest include economics, labour and finance; environment; public governance; science and technology; tax policy; trade and agriculture; energy, transportation and sustainable development; and education, which is the focus of this document.

The OECD currently consists of 34 countries,[3] providing a budget of EUR 363 million in 2015 (OECD, 2015a). Each country contributes in proportion to its GDP; the top five contributors in 2015 were the US, Japan, Germany, France and the UK. Like other IOs, the OECD is 'simultaneously, a geographic entity, an organisational structure, a policy-making forum, a network of policy makers, researchers and consultants, and a sphere of influence' (Henry, Lingard, Rizvi, & Taylor, 2001, p. 7), but internally it is articulated around three main bodies:

1) a Council, which is 'the body from which all acts of the Organisation derive' ("Convention on the Organisation for Economic Co-operation and Development," 1960, Art. 7);
2) a 'complex, hierarchical committee system' (Carroll & Kellow, 2011, p. 9), composed of 'about 250 committees, working groups and expert groups' (OECD, 2015j) listed in the "On-Line Guide to OECD Intergovernmental Activity" (OECD, 2015g, hereinafter the "Guide to the OECD");
3) and the Secretariat, the sole body to employ staff internal to the organisation, whose role is to 'support the activities of committees, and carry out the work in response to priorities decided by the OECD Council' (OECD, 2015j).

As is the case for other organisations, the three systems are highly interdependent and the internal structure of the organisation is difficult to pin down: committees and sub-committees come into existence for the lifetime of a project and then disappear; staff enter, exit and move within the organisation; restructuring can be structural or merely cosmetic.

Institutional theories have a long history in sociology, economics and political sciences (Scott, 1995), whereas organisations and their epistemology have been the subject of numerous 'paradigm wars' (Moldoveanu & Baum, 2005) between realists and constructivists. Because of their complexity, large-scale organisations have been likened to machines,

[3] In alphabetical order (with date of accession): Australia (1971), Austria (1961), Belgium (1961), Canada (1961), Chile (2010), Czech Republic (1995), Denmark (1961), Estonia (2010), Finland (1969), France (1961), Germany (1961), Greece (1961), Hungary (1996), Iceland (1961), Ireland (1961), Israel (2010), Italy (1961), Japan (1964), Korea (1996), Luxembourg (1961), Mexico (1994), Netherlands (1961), New Zealand (1973), Norway (1961), Poland (1996), Portugal (1961), Slovak Republic (2000), Slovenia (2010), Spain (1961), Sweden (1961), Switzerland (1961), Turkey (1961), United Kingdom (1961), United States (1961).

organisms, brains, cultures, political systems, instruments of domination or even psychic prisons (G. Morgan, 2006).

Such a wealth of literature, attempting to explain why organisations are the way they are, is for the most part only tangential to this research. However, the underlying idea that international organisations are complex phenomena (i.e., complicated and irreducible to some components, see Rogers, 2008) with some emergent properties—such as a certain degree of agency and autonomy from member states—underpins the whole study and will be invoked to clarify some apparent contradictions in the OECD's "behaviour".

Why is the OECD important for education?

Education and economics: 1961–1988

The OECD was created as an economic platform, and indeed in the 1960s education was at the margin of the OECD's work and at the service of the economic agenda. In line with Human Capital theory, investment in education was seen as 'a precondition for producing an adequate supply of scientists and technicians' (Martens & Jakobi, 2010, p. 164) in order to win the space and big science race (Rinne, Kallo, & Hokka, 2004). Human capital encompasses a person's 'knowledge, skills, health, or values' (Becker, 1964/1993, p. 16), as opposed to his or her physical or financial assets. Education is presumed to foster human capital which, in turn, should promote economic growth (Romer, 1990).

A noticeable change in the status of education within the OECD was observed with 'the establishment in 1968 of the Centre for Educational Research and Innovation (CERI, initially with Ford Foundation and Shell funding), and the replacement of the Committee for Scientific and Technical Personnel [which until then had dealt with education matters] with the Education Committee in 1970' (Henry et al., 2001, p. 9).

In the late 1960s and early 1970s, education in the OECD acquired a strong social democratic component. The reductionist view of education as a means to increase the stock of human capital was challenged by a strong Secretariat and by country-level trust in Keynesian economics, which gave governments a more prominent role in public policy. In this period 'the "social" came to dominate the "economic"' (Henry et al., 2001, p. 64; quoting Papadopoulos, 1995). The focus was on "recurrent" education, a precursor of what today is known as "lifelong learning", and its role in social emancipation and equality (Bengtsson, 2008; Martens & Jakobi, 2010).

From the mid-late 1970s into the 1980s, neo-liberalism and its free-market ideology replaced Keynesian economics in many OECD countries and sought to minimise government interventionism (Carroll & Kellow, 2011; Tony Porter & Webb, 2007). Following the failure

of the Bretton Woods system of economic policies and the 1973/74 oil crisis (see Eichengreen & Kenen, 1994, for an overview), the question of whether education policies were the best instrument for addressing social inequalities came to the surface (Blaug, 1976), and education was once again placed at the service of economic growth (Papadopoulos, 1995). Concepts such as "adaptable workforce" emerged (Henry et al., 2001; Papadopoulos, 1995) whereas education reforms were framed as part of an integrated political response to the needs of a changing society.

Education in the spotlight: the INES programme

Over time, the role of the OECD had sharply declined in financial and monetary matters but had acquired new prominence as a source of evidence-based analyses, especially in the trade sector (Henry et al., 2001). In education, the demand for comparative (and comparable) data had increased (Henry et al., 2001; C. Morgan, 2011; von Bogdandy & Goldmann, 2012). At the time, the OECD did not have the capacity to produce its own indicators and had to rely on country-reported statistics that were often outdated and hardly comparable (Martens & Jakobi, 2010). Earlier attempts to relate education systems had revealed the inadequacy of the theoretical models used to account for country differences (Henry et al., 2001; Rinne et al., 2004). As a result, 'a culture of distrust towards performance indicators had grown up over the years' within the CERI (Henry et al., 2001, p. 87).

However, calls for education reforms within influential countries such as the US (National Commission on Excellence in Education, 1983) and France resulted in high political pressure on the OECD to develop effectiveness (input and output) indicators. Eventually, the CERI 'had no choice but to concede' (Henry et al., 2001, p. 87). In 1988, the International Indicators and Evaluation of Educational Systems (lNES) project was born, and in spite of the attempt by the CERI to underplay—and to an extent undermine[4]—its success, it rapidly became tremendously popular among member countries (Henry et al., 2001). Moreover, the timing of the programme coincided with the crumbling of the USSR, a wider OECD strategy to reach out to non-Western countries and collaborations with UNESCO and Eurostat on the development of further indicators. All this translated into a considerable expansion of the OECD's area of influence on educational matters (Henry et al., 2001; Rinne et al., 2004).

The work on indicators was carried out by networks led by one country at a time and assisted by a scientific advisory group and the Secretariat (Henry et al., 2001). Among these, 'Network A, which was charged with the development of indicators on the outcomes of

[4] 'CERI's position on indicators was reflected in the decision to place the INES project in charge of a member of the secretariat with a philosophical background in the expectation, perhaps, that it would not succeed' (Henry et al., 2001, p. 88).

education, acquired particular prominence' (von Bogdandy & Goldmann, 2012, p. 56). In 1992 the work on the INES project, and especially of Network A, flowed into the 'flagship' publication (C. Morgan, 2011, p. 54; von Bogdandy & Goldmann, 2012, p. 56) of the OECD in education, the *Education at a Glance* (EAG) series, comprising an array of dozens of indicators including demographic characteristics, resources, access to education, learning environments and outcomes.

The scope and purposes of the EAG series were the object of great internal debate and refinement. The core assumption behind the newly-developed statistical indicators was that national practices differed but the underlying constructs and conceptions of education were the same across all countries. This assumption was overly optimistic. Where generic indicators such as "lifelong learning", "equity" or "quality of provision" created 'categories of equivalence […] to give [the illusion of] uniformity among diversity' (Popkewitz, 2011, p. 31), the OECD was struggling to reconcile the neo-liberal and neo-conservative views of some of its members with more egalitarian ones (Henry et al., 2001).

The former promoted decentralization and accountability, a market/business-like conception of education and individualistic views of equity. The latter, instead, saw education as a vehicle for social cohesion and were backed by the CERI, which had not given up on the idea of developing better theories before placing confidence into indicators. Specifically, the CERI was interested in how the emerging Social Capital Theory (J. S. Coleman, 1988)—"social capital" was broadly speaking the value of social relations—could fit within education. The interaction of all these competing views meant that for years the OECD conveyed mixed messages about what equity, quality and efficiency actually meant (Henry et al., 2001).

Emancipation from external data sources: the development of PISA

For the publication of EAG, Network A assembled indicators produced by existing international studies (C. Morgan, 2011), such as those of the International Association for the Evaluation of Educational Achievement (IEA) and the International Adult Literacy Survey (IALS) of 1994, which was 'a collaborative effort involving several international organizations, intergovernmental agencies, and national governments' (Kirsch, 2001, p. 1). There was growing discontent with this practice because the OECD did not have control over costs, timing, assessment contents and procedures, or even which countries to involve (C. Morgan & Shahjahan, 2014; citing Owen, Hodgkinson, & Tuijnman, 1995).

Therefore, in 1995, in a period in which the organisation was 'searching for relevance in a changing world' (Carroll & Kellow, 2011, p. 114), Network A presented a strategy to the other members of the INES project to develop their own assessment (C. Morgan, 2011). The strategy was accepted by the Council which, in 1997, authorised the Secretary-General 'to

implement a decentralised Programme for Producing Student Achievement Indicators on a Regular Basis' (OECD, 1997d, p. 2). With this move, the OECD 'transformed itself from a gatherer of educational indicators to a producer of educational data' (C. Morgan & Shahjahan, 2014, p. 196). Between the third and the fifth meeting of the Board of Participating Countries (now PISA Governing Board, PGB), the official name of the enterprise became Programme for International Student Assessment, or PISA (OECD, 1998b, 1998c).

The importance education had acquired within the OECD thanks to the INES programme, EAG and PISA, was further increased by the link between education and growth that a new economic paradigm—knowledge-based economy, 'one in which the generation and the exploitation of knowledge has come to play the predominant part in the creation of wealth' (DTI, 1998; cited in Houghton & Sheehan, 2000, p. 1)—made possible (OECD, 1996). Knowledge economy provided a new identity to the OECD (and other IOs), one based on four 'pillars' (Robertson et al., 2007): information technology, openness to international trade, national innovation systems and education and training for lifelong learning.

Thus, at the end of the 1990s, the OECD became an influential actor in educational matters by working on two interconnected fronts. On the one hand, by subscribing to the narrative of knowledge economy, it invested in a certain type of capital; on the other hand, by taking control of educational monitoring, it became the official source for the allocation of such capital to each of its members. 'Literacy and numeracy skills have become a currency in modern societies around the world' (OECD, 2010g, p. 33); the OECD is 'involved in helping to specify the skills and competencies that give contemporary human capital its value' (Sellar & Lingard, 2013b, p. 718).

What is PISA?

PISA was established as a three year assessment (OECD, 1997d), and this scheduling never changed. Thus far, it has been run in 2000, 2003, 2006, 2009, 2012 and 2015, although data for this latter cycle were not available at the time of writing.

In the educational literature PISA is categorised as an "international assessment", "large-scale assessment" or "cross-country survey". From an analytical perspective, it is a time-series cross-sectional survey (Fairbrother, 2013; Nagel et al., 2010), because it follows countries over time (the 'time-series' component) while sampling a different group of students at each cycle (the 'cross-sectional' one). In this document, PISA and similar surveys—such as the Trends in International Mathematics and Science Study (TIMSS) and the Progress in International Reading Literacy Study (PIRLS), both developed by the IEA—are referred to as "large-scale assessments" (LSAs).

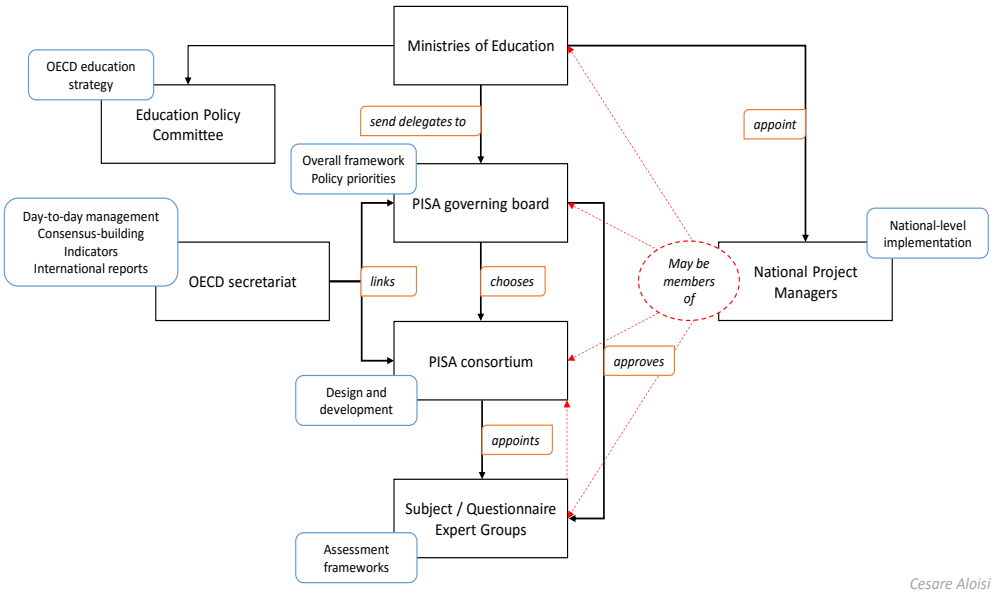
PISA has, up to know, been generally a paper-and-pencil assessment, although countries also have the option to participate in a digital version. The test lasts two hours for each student. It uses a combination of multiple-choice, short answer and extended response items and all items are distributed in 8–9 different but linked booklets, each answered by one student (Adams & Wu, 2002). Participating students are sampled by age within a specific window: eligible students must be aged from 15 years and 3 months to 16 years and 2 months at the beginning of the assessment period, plus or minus one month, and be enrolled at least in Grade 7 (OECD, 2005b, p. 46).

Rather than focusing exclusively on academic competence, PISA takes a broader perspective and claims to be measuring 'skills for life' (OECD, 2001a). These skills are expressed in terms of three core literacies (though additional domains such as problem solving and financial literacy have also been assessed): Reading of the home language, Mathematics and Science[5]. PISA uses items that are cross-nationally agreed upon but as far as possible context- and curriculum-free (Adams & Wu, 2002; OECD, 1999a, 2001a). This approach was strongly advocated by the Australian Council for Educational Research (ACER, see OECD, 1998c), which won the first call for tender to develop the instrument (OECD, 1997b) and still has a leading role (see for example OECD, 2014h, p. 462).

PISA is currently developed and administered by the network in Figure 1.1.

[5] The OECD "Reading" is defined as: 'understanding, using and reflecting on written texts, in order to achieve one's goals, to develop one's knowledge and potential, and to participate in society' (OECD, 1999a, p. 12). "Mathematics" is defined as 'Identifying, understanding and engaging in mathematics and making well-founded judgements about the role that mathematics plays, as needed for an individual's current and future life as a constructive, concerned and reflective citizen'. "Science" is defined as 'Combining scientific knowledge with the drawing of evidence-based conclusions and developing hypotheses in order to understand and help make decisions about the natural world and the changes made to it through human activity' (*ibid*.).

Figure 1.1: The organisational structure behind PISA



Source: own elaboration

The PISA Governing Board (PGB)

> 'Countries appoint representatives to the Board who are knowledgeable about large-scale student assessments and their interface with educational policy and practice. Representatives comprise a mix of government officials and staff of research and academic institutions. […] The Board determines the policy priorities for PISA and oversees adherence to these priorities during implementation. This includes the setting of priorities and standards for data development, analysis and reporting as well as the determination of the scope of work that will then form the basis for the implementation of PISA. Economies that participate in PISA but do not have associate status are welcome to participate in PGB meetings as observers.' (OECD, 2015h)

The PGB has also the task of 'disseminat[ing] its policy advice, analysis, research and data to a wide range of Member and non-Member stakeholders' (*Guide to the OECD*: PGB). The PGB 'seeks the guidance of the Education Policy Committee' (*ibid.*), which in turn is responsible for 'assisting governments to develop effective, efficient and evidence-based policies for education and learning to meet individual, social, cultural and economic objectives through the development of specific policy recommendations, policy reviews, analyses and data collection' (*Guide to the OECD*: Education Policy Committee).

The PISA Consortium

'The design and implementation of the surveys, within the framework established by the PISA Governing Board, is the responsibility of external contractors under the leadership of the

ACER' (OECD, 2014h, p. 462). These contractors are chosen by the PGB 'through a transparent and open competitive process' (PISA Governing Board, 2014).

The Expert Groups

Expert Groups are in charge of advising the Consortium on the development of the surveys for Reading, Mathematics, Science, Problem Solving, Financial literacy and the background questionnaires. These groups are 'appointed by the main PISA contractors with the approval of the PISA Governing Board' and they include subject experts as well as 'experts in assessment, technology, and education research from a range of countries' (Mathematics Expert Group, 2010, p. 4). In fact, if one considers the core domains and the background questionnaires, 24 countries are represented within the Expert Groups, but only 10 of them (Australia, Belgium, Denmark, Finland, Germany, Netherlands, Norway, Poland, the UK and the US) had a national expert in at least 4 out of 5 (2000–2012) PISA administrations (see Appendix 3, Table 1.A, for a list).

The PGB has often asked the Consortium to 'address the imbalance towards North American and European expertise when positions in the groups become vacant' (OECD, 2010a, p. 4), but argues that 'the limited size of the expert groups does not mean that only a limited number of countries will participate in the development of the project' (OECD, 2005a, p. 22). The Expert Groups certainly consult advisors from more than 10 countries (OECD, 2014h), but the extent to which suggestions are eventually integrated into the assessment is not clear from OECD reports.

National Project Managers (NPMs)

NPMs are individuals or organisations in charge of implementing the survey at the national level following standardised procedures. Other responsibilities include processing data and disseminating results at the national level by organising conferences, seminars and by preparing country-level reports. Country reports are needed to 'contextualise the findings in the international reports[6] through national analysis and interpretation as determined by each country in ways that respond to the specific policy and educational context' (OECD, 1999c, p. 4).

The relationship between NPMs and the other groups, represented by the dashed red arrows in Figure 1.1, shows that there are no clear boundaries separating an international (or external) level in the development of PISA and a national or internal one. As leading research centres or prominent academics, members of the Consortium and the Expert Groups are likely

[6] The "international reports" are the OECD publications summarising the outcomes of each administration of PISA, such as *PISA 2012 Results: What Students Know and Can Do* (OECD, 2014g), and available at http://www.oecd.org/pisa/pisaproducts/.

to have ties with national-level policymaking or hold posts of responsibility. Similarly, NPMs can sit in more than one committee.

For instance, the ACER is in the PISA Consortium and in charge of the Mathematics framework, while ACER staff are NPMs for Australia and members of the Technical Advisory Group (OECD, 2014h). Likewise, the service for the Analysis of Systems and Practices in Education (ASPE) is in charge of administering PISA in the French Community of Belgium, but it is also represented in the PISA Governing Board, the PISA consortium, the Technical Advisory Group and the Reading Expert Group.

Although some countries are more involved than others in the development of PISA (the United States, Australia, Japan, Germany, Belgium, the Netherlands), in general in only a minority of countries' NPMs are *not* part of the PGB. Table 1.B in the Appendix shows that, in 2015, in 72% of cases NPMs and representatives at the PGB came from the same institute or were actually the same person. Overall, this suggests that a clear separation between the "international" and the "national" may exist only between the OECD Secretariat—composed of international civil servants, who are 'citizens of OECD member countries […] with no national affiliation during their OECD posting' (OECD, 2008b, p. 12)—and everyone else.

The Secretariat

The Secretariat has an administrative side, which is at the service of OECD countries and constrains its operations; but also a scientific one, as a network of researchers enjoying some academic and managerial freedom. For instance, they can collaborate with external consultants whose views may differ from those of the governments and reinforce the Secretariat's voice on policy issues (Henry et al., 2001). The pull between dependence and independence is a consequence of the historical development of the predecessor of the OECD, the OEEC. The OEEC was forced to compromise between the positions of countries that thought the Secretariat should be under the strict control of member states, and those that saw a unique opportunity in granting it more independence (Carroll & Kellow, 2011).

Within the Secretariat at the OECD, the Directorate for Education and Skills works under the supervision of the Education Policy Committee and in collaboration with the CERI, the PGB and the Institutional Management in Higher Education Governing Board. It branches into four divisions: Early Childhood and Schools (managing PISA), Innovation and Measuring Progress, Policy Advice and Implementation and Skills Beyond School (OECD, 2015f).

Some researchers have argued that Secretariats at international organisations have a considerable power to shape policies—especially when their staff is 'highly qualified, argumentative and mission-driven' (Yi-Chong & Weller, 2008, p. 35). Yi-Chong & Weller (2008) examined the role of international civil servants in the Secretariat of the World Trade

Organization. They traced their influence to the fact that, even if they do not have any last word or executive power over a matter, they have creativity, 'technical expertise and bureaucratic skill' (p. 37), control over information, sometimes a position with recognised leadership, and the professional longevity to 'outlast' the experiences 'of most state delegates' (p. 40).

Within the principal-agent theory framework (Eisenhardt, 1989, for an overview), delegated agents such as the Secretariat are 'actors in their own right' (Hawkins, Lake, Nielson, & Tierney, 2006, p. 5), and Shober (2010) argues that, even when they already enjoy some discretion, 'with a particular combination of factors, [agents] are able to […] shape the preferences (i.e., policy goals) of their principals' by leveraging their 'already-granted scope to gain more autonomy' (p. 25, emphasis removed).

This can happen, for instance, under a strong and charismatic leadership (Shober, 2010). In the case of the OECD, a current key figure in the Secretariat and PISA is the Director for Education and Skills—and Special Advisor on Education Policy to the Secretary-General—Andreas Schleicher. Schleicher was previously professionally linked to the IEA and CERI. His dominant role in the success of PISA emerged clearly during the field work carried out for this document. Although the survey is not a "one-man endeavour", the OECD Secretariat acquired a considerable influence in educational matters under his management: 'Andreas Schleicher travels the world proselytising PISA and has been very successful' (Grek, 2012, p. 248).

Moreover, the Secretariat at the OECD has an active role in peer reviews—a "signature" activity of the organisation (see "What motivates countries to participate to PISA?", below). Even though they are carried out by member countries, the Secretariat 'supervises the whole review process by providing documentation and analysis, […] often suggests a country for review, selects the reviewer, makes personal visits during the examination, and instructs its staff to prepare the publications' (Martens & Jakobi, 2010, p. 11). Given its central role in handling knowledge, the Secretariat has the potential to exert substantial influence on other bodies such as the Council, the Committees or, in the case of PISA, the PGB.

Nevertheless, the power held by the Secretariat or the Directorate for Education and Skills should not be overstated, because their autonomy is 'constrained' (Carroll & Kellow, 2011, p. 42) by the interests of all but hands-off stakeholders. Representatives from member countries are 'deeply engaged' (p. 29) with the work of the Secretariat and the Directorates—and a reminder of this is the country presence in all key positions highlighted by Figure 1.1.

Evaluation mechanisms are in place, and each policy draft produced is the result 'of a complex process of communication, negotiation, bargaining and co-ordination between internal and external stakeholders' (Carroll & Kellow, 2011, p. 42). Since the beginning of the

PISA enterprise, the PGB has often ‘asked’ (OECD, 1998a), ‘requested’ (OECD, 1999b), ‘recommended’ (OECD, 1999c) or ‘reiterated’ (OECD, 2001c) that a series of actions should be taken. It has expressed ‘disappointment’ (OECD, 2001c) about actions of the Directorate at times even ‘concerns’, for instance, ‘about the quality of the draft’ or ‘the nature and language of the report [, which] was not sufficiently oriented to policy analysts and policy makers as the main intended audience’ (OECD, 2002b, p. 6).

Kellow & Carroll (2013) note that the power of the Secretariat at the OECD ‘relies primarily upon its ability to persuade [...] members of the accuracy and value of its data, its reports and the conclusions and recommendations they contain’ (Carroll & Kellow, 2011, p. 29). However, at the turn of the century it came under considerable criticism because of a lack of transparency in its career structure and appointment practices, with a widespread tendency among line managers to employ highly specialised short-term staff for specific projects without any formal recruitment process (Carroll & Kellow, 2011). This led to a significant restructuring of human resourcing and process management between 2004 and 2007.

This is why it would be inaccurate to refer to the Secretariat or the Directorate as “bureaucracies”. A bureaucracy is a rule-creating institution characterised by a stringent division of labour, career longevity, impersonality and independence from politicisation, and expertise—especially control over information (M. N. Barnett & Finnemore, 1999, 2004; for an original definition, see for example Weber’s essays in Gerth & Wright Mills, 1946). Such control endows the bureaucracy with authority: ‘the ability of one actor to use institutional and discursive resources to induce deference from others’ (M. N. Barnett & Finnemore, 2004, p. 5).

It can certainly be claimed that the authority of the Directorate for Education and Skills stems from its perceived—or at least self-proclaimed—impartiality and expertise, as well as its ability to classify the world, fix meanings and diffuse norms (M. N. Barnett & Finnemore, 2004). However, it is also a body characterised by short careers, lack of absolute ownership in the production of information, as well as by a persistent tension between the creative, scientific and epistemic work expected by the highly qualified staff employed at the organisation, and the reality of having to satisfy ‘externally defined national interests for utility maximization’ (Marcussen & Trondal, 2011, p. 611).

What is the purpose of PISA?

PISA was conceived primarily as tool to inform policymaking. Within the broader OECD framework, PISA ‘reflects and complements the priorities of the OECD Members, in promoting sustainable development and social cohesion through good governance’ [*Guide to the OECD*: PGB]. Thanks to ‘trend indicators that allow participants to monitor improvements

in educational outcomes', PISA 'seeks to progressively enhance comparative policy insights on how to improve: the quality of learning outcomes; equity in learning opportunities; the effectiveness and efficiency of educational processes; and the impact of learning outcomes' [*Guide to the OECD*: PGB].

The international reports were meant to be complemented by thematic reports taking 'a policy-oriented perspective' (OECD, 1999b, p. 3) and ensuring PISA outcomes could be situated 'in the context of other qualitative and quantitative work at the OECD' (*ibid*.). In general, thematic reports were to focus on institutional characteristics with 'policy analysts and policy makers as the main intended audience' (OECD, 2002c, p. 5).

In November 2002, the first international PISA symposium was held. The nature of the symposium was political, not academic, and its primary purpose was to discuss the impact of PISA on policies and draw lessons from systemic change (OECD, 2002a, 2002b). Education ministers at the OECD saw the potential of PISA for the medium and long term, and suggested that one of the aims should be to produce 'clear signals that can inform education policy on how to improve the effectiveness of education' (OECD, 2001c, p. 4).

PISA's objectives were to improve the quality, equity and efficiency of education systems, and to increase the '"durability" of knowledge and skills and their social and economic relevance' (OECD, 2003b, p. 11)—the 'skills for life' that were promoted ever since. To reach these objectives, each administration cycle had to emphasise different policy themes (OECD, 2002b), and 'policy levers' (i.e., possible avenues for policy action) had to be established (OECD, 2003b, p. 11).

The overall strategy for PISA did not change substantively over time: in 2013 the PGB reiterated that PISA 'is *policy-oriented*'; it provides 'data and analysis that can help guide decisions on education policy' and 'identifies features common to high-performing students, schools and education systems' (OECD, 2013a, p. 3, emphasis in the text). Martens & Jakobi (2010) argued that, with PISA, the OECD can explore topics that are relevant for education policy and 'link them immediately to issues of policy formulation and implementation' (p. 175).

The clear policy orientation of PISA was not matched by a definite strategy in terms of what education policy areas to focus on, how to translate PISA results into policy advice and how recommendations should be communicated.

With regard to the first point, the core literacies of Reading, Mathematics and Science were an easy choice as they provided an opportunity 'for innovating and informing national curriculum development' in these areas (OECD, 1998b, p. 6). Additional policy areas were identified as possible candidates for PISA-informed thematic reports: equity issues, student and school factors, parental involvement, gender differences, streaming, but also 'policy

options for school improvement in cross-curricular and non-curricular areas' (OECD, 1999c, p. 5). The main instrument for collecting data on these topics was the set of contextual questionnaires, whose purpose was to 'contribute to indicators of enduring cross-national relevance […and] focus on areas that are amenable to policy development' (OECD, 1999c, p. 7).

The second point relates to the transition from PISA data to their interpretation (i.e., the knowledge-construction process). This is a key aspect as it is linked to the validity of the assessment. The Directorate for Education and Skills was charged with interpreting PISA data but there is evidence, which is reported in a later section, that there was no explicit protocol guiding this activity.

With respect to the third point, a dissemination strategy was agreed between Schleicher and the PGB. The international reports had to be 'as country specific as necessary to be interpretable in the national institutional context […and] as accessible as possible for a policy audience', while still remaining comparable and valid cross-nationally (OECD, 2001c, p. 6). The PGB asked the Secretariat 'to ensure a balanced representation of all countries and to avoid focusing on extremes only' (OECD, 2001c, p. 6), and has at times requested to remove 'normative and speculative recommendations to individual countries' (OECD, 2003b, p. 5).

Whether this goal was achieved is up for scrutiny. The OECD aims to be seen as an independent expert, an external objective advisor reporting scientific facts. But it is also a "critical friend", one whose reports and analyses 'are subtly normative' (Henry et al., 2001, p. 14): countries are often singled out as evidence backing a policy point, or as a way to provide a success story or an exemplar case. Detractors have claimed that PISA reports use a tactic of 'naming and shaming' (Armingeon, 2004; Carvalho & Costa, 2014), which the OECD already used in domains other than education to exert influence on member countries.

The practice of ranking countries in international reports (established in fact by earlier IEA studies) has also been questioned because of its limited usefulness for policymaking (Goldstein, 2004; Grek, 2009). In this respect, the official PGB position has been both that reports 'should highlight important policy messages emerging from PISA and therefore go beyond the ranking of countries' (OECD, 1999c, p. 3), and that rankings have their place if appropriately contextualised (OECD, 2001b). Overall, the evidence suggests that PISA is a policy tool supported by empirical evidence, rather than a scientific measurement instrument *per se*.

How is PISA used to inform policymaking? Theory and evidence

What motivates countries to participate in PISA?

Participation in PISA has been consistently strong. Thirty two 'economies' or 'jurisdictions' (this includes countries, disputed territories and, more recently, cities) participated in the first administration in 2000; eleven countries followed in 2002 (OECD & UNESCO-UIS, 2003a), to make a total of 43 participating in the first cycle. After a slight decrease to 41 jurisdictions in 2003, the number has been on the rise: 57 in 2006, 65 in 2009—with 9 additional countries administering the 2009 assessment in 2010—65 in 2012; 71 'are signed up to participate in 2015' (OECD, 2015h). This is an impressive achievement, considering that the OECD managed in 15 years to create a participant base as large as IEA's, which has been running large-scale assessments since the 1960s.

The attraction of PISA can be explained by at least three factors (for a further level of investigation, see DeBoer, 2010). Firstly, PISA data are relevant for (established or aspiring) knowledge-based economies because they are a more accurate estimate 'of the stock of human capital or human skills' (OECD, 1999a, p. 11) than other commonly-used measures (e.g., highest attained qualification). Secondly, the assessment sits comfortably within the lifelong learning framework for education of both the OECD and the European Union (CERI, 2008; Eurydice, 2001; Saar, Ure, & Holford, 2013). Thirdly, PISA is a perfect instrument for the OECD's governance model. This third aspect is now analysed more in detail.

According to the principal-agent theory, the OECD is a "collaboration agent"—a platform that facilitates collaborative policymaking by collecting and revealing information about country efforts to tackle problems that are collectively perceived as shared (Hawkins et al., 2006). The OECD describes itself as 'a group of like-minded countries' (OECD, 2008b, p. 8), and others have defined it 'a comparative forum, accommodating both a sense of national autonomy as well as a sense of commonality among the like-minded' (Henry et al., 2001, p. 56).

The idea is that individual members (principals) delegate some authority to a superstructure (the OECD, the agent) to carry out certain tasks—namely, providing or enabling feedback about individual country policies. A distinguishing feature of the OECD lies in the method of choice that has traditionally been employed to collect country data and provide such feedback: peer reviewing, whereby one country's practices in a certain area are examined by other members.

Peer reviews are a way to share best practices but also to put countries under pressure by highlighting their shortcomings, in hopes to achieve a certain degree of policy transfer, adjustment and 'voluntary' convergence (Bieber & Martens, 2011). Peer pressure 'can act as

a powerful incentive to improve policy' (OECD, 2008b, p. 7), especially when it turns into 'naming and shaming', which happens when countries perpetuating some harmful practices (for instance, failure to introduce anti-money-laundering legislation) are singled out (Carroll & Kellow, 2011).

It is not that the OECD lacks legal power; for instance, it can make Decisions, which are formal acts that are 'legally binding on all those Member countries which do not abstain at the time they are adopted' (OECD, 2015e). However, these instruments are seldom used, and the organisation prefers to govern through 'soft mechanisms' such as persuasion, surveillance and self-regulation (Martens & Jakobi, 2010, p. 7):

> 'The OECD achieves results mainly by convincing governments of the value of adopting policies that are in their collective self-interest. It does not disburse grant money and it has no disciplinary powers other than moral [per]suasion.' (OECD, 2011a, p. 12)

PISA fits well within the 'hortatory nature' (Carroll & Kellow, 2011, p. 34) of the OECD authority because of its quantitative approach. Nowadays, 'peer' comparisons as a mode of governance to regulate country behaviour are often carried out quantitatively through statistical reports. This is why "the politics of mutual accountability" (Nóvoa & Yariv-Mashal, 2003), and "governance by comparison" (Martens & Niemann, 2010), have also been termed "governance by numbers" (Grek, 2009), "the politics of/as numbers" (Lingard, 2011) and "the tyranny of numbers" (Ball, 2015).

The role of numbers in country politics had already been shown by Desrosières (2002). The general argument is that governance by numbers works by decontextualising differences—or recontextualising them 'into more global comparative frameworks' (Henry et al., 2001, p. 104)— to make them comparable. Numbers are 'strategies of communication' in that they 'conveniently summarize a multitude of complex events and transactions', and are therefore 'well suited for communication that goes beyond the boundaries of locality and community' (Theodore Porter, 1995, pp. viii–ix).

PISA produces a highly numeric kind of knowledge. Students' ability in Reading, Mathematics and Science is measured and scaled, and their overall proficiency is assigned a numbered level. Students' estimated scores are aggregated and countries ranked on a scale with a mean of 500 and a standard deviation of 100: this is the number that purportedly defines how good overall an education system is in getting young people ready for the 21st century and which captures the attention of policymakers and the media.

Even more important than country outcomes, from a policymaker's perspective, are country inputs and processes—i.e., the features successful countries have and which might explain the observed scores. In this, a key role is played by statistical indicators such as

"GDP", "SES" or "school autonomy", whose correlations with country scores are extensively analysed in international reports.

Born out of the neo-positivist Social Indicators Movement in the 1960s (D. Rutkowski, 2008; Smith & Baker, 2001), educational indicators are used as if they were neutral quantitative measures, but the reality is more contested. Indicators capture a certain way of organising the world and as such they are charged with political meanings and objectives. As the OECD states: indicators 'must be able to communicate a story to decision-makers and other end-users quickly and accurately' (OECD & Joint Research Centre of the European Commission, 2008, p. 40). Take, for example, the use of indicators of The World Bank:

> 'The World Bank does not only collect data and produce descriptive statistics on national economies. It also takes those raw data and couples them to particular policy problems, often of the Bank's own creation. The Bank defines development, telling us what data measure it. It tells us what constitutes poverty and what data are necessary to act on that policy problem.' (M. N. Barnett & Finnemore, 2004, p. 7)

Thanks to PISA, in a 'self-perpetuating dynamic' (Sellar & Lingard, 2013b, p. 722), the OECD defines an educational issue, generates data and analyses them (Martens & Jakobi, 2010) at a time 'when data have become central to the new governance at both global and national level' (Sellar & Lingard, 2013b, p. 716). This makes it 'an international mediator of knowledge' (Henry et al., 2001, p. 84), an 'artist', arbitrator and authority in the diffusion of ideas (Marcussen, 2004)—but also a supplier of the new 'currency' embodied by the three literacies (OECD, 2010g, p. 33). It is what Sellar & Lingard (2013b) call "epistemological governance". The OECD situates itself at the centre of an epistemic community—'a network of professionals with recognized expertise and competence in a particular domain and an authoritative claim to policy-relevant knowledge within that domain or issue-area' (Haas, 1992, p. 3).

Far from being a neutral exercise, PISA defines what reality to look at and what it should look like. It frames an educational realty, but also a moral one, 'a single comparative field which pivots around certain normative assumptions about provision and performance' (Henry et al., 2001, p. 84); it is in this field that countries assess their strengths and weaknesses against each other. Woodward (2009) speaks of a "cognitive" model of governance, which fixes meanings, and a "normative" one, which spreads norms about what constitutes good (educational) governance (M. N. Barnett & Finnemore, 2004; see also Jones, 2007, on "norm entrepreneurs").

Beyond PISA, Marcussen (2004) argues that the OECD formulates, transfers and authorises both causal beliefs, which 'can range from theories on the one hand, to facts and data on the other', and principled beliefs, which 'can be anything from visions and good

practices to norms for appropriate behaviour' (p. 16). Therefore, the OECD governance model is as much about knowledge as it is about identity (Tony Porter & Webb, 2007).

To answer the title question (What motivates countries to participate to PISA?), one could say that the OECD is "preaching to the converted", and add three points. Firstly, the OECD's role in the conversion is supported by overarching paradigms dating back to the 1960s, which link education to the economy through the conceptual frameworks of human and social capital theory, knowledge-based economy and literacy as a skill. Secondly, participating countries willingly subscribe to the governance by comparison model: in other words, they want to be converted. Standing on the PISA scales gives countries, particularly those outside the OECD, a level of exposure in educational matters which they might not otherwise attain. And finally, the incredible work done by Andreas Schleicher should not be underestimated. It may be a coincidence, but after he became Special Advisor to the Secretary General, PISA featured at the very beginning of the OECD's 50th anniversary report, even before the organisation's efforts to combat climate change (OECD, 2011a, p. 5).

How has PISA been received and used for national policymaking?

It has been argued theoretically (Dale, 1999, 2000) and shown empirically (L. Rutkowski & Rutkowski, 2009) that countries respond differently to globalising forces. In many cases, PISA results seem to be exploited to initiate or justify national policies that are somewhat disconnected to international recommendations (Figazzolo, 2009; Grek, 2009), yet a whole spectrum of country responses to international data has been reported (Carvalho & Costa, 2014; Knodel, Martens, de Olano, & Popp, 2010). This includes extreme reactions such as self-scandalisation, self-glorification (Steiner-Khamsi, 2004a) or a 'sabre-rattling political rhetoric to drive through educational reforms' (Baird et al., 2011, p. 2); but also indifference ("the most frequent policy response", according to Steiner-Khamsi, 2004a, p. 208) and reflective learning approaches (Baird et al., 2011).

In this section, specific country reactions to PISA are summarised to understand whether—in spite of its policy-driven intentions—the survey ended up being merely an exercise in international ranking for political accountability, or whether the OECD successfully managed to create a "public policy instrument", 'a device that is both technical and social, […] carrying a concrete concept of the politics/society relationship and sustained by a concept of regulation' (Lascoumes & Le Galès, 2007, p. 4).

In Canada and in England, PISA is used for public accountability (Grek et al., 2009; Volante & Ben Jaafar, 2008). The former Education Secretary, Michael Gove, stated that Schleicher is 'the most important man in English education' (Exley, 2013) and 'the father of more revolutions than any German since Karl Marx':

> ‘Because Andreas is responsible for collating the PISA league tables of international educational achievement. He tells us which nations have the best-performing education systems and then analyses that data to determine why that is the case.’ (Gove, 2012)

Public accountability can be seen in Macao, where PISA has informed national and school assessment policies and is being used by the central government to evaluate the effectiveness of quality assurance interventions (Vong & Wong, 2014). A well-known and often cited reaction was the ‘PISA-shock’ in Germany, which is covered later on in this section. A less-well documented PISA shock was experienced by Norway, and it provided an opportunity for the newly-established government to legitimise a series of curricular and assessment reforms which were rapidly implemented (Baird et al., 2011; Tveit, 2014).

Looking at several case studies, it is difficult to discern when new policies were informed by evidence from PISA and when instead reformists appropriated a certain language or appealed to an external authority to push previously-designed interventions.

In New Zealand, an educational ‘task force’ was established to evaluate the implementation of a new curriculum that had already been designed (Baker & Jones, 2005). Baird et al. (2011) report that, in Switzerland, some policy interventions following PISA findings ‘had been brewing for decades’ (p. 32). PISA helped only to legitimise the arguments of reform proponents (Bieber & Martens, 2011).

Pons (2012) analysed the reception of PISA in six countries (France, the French Community of Belgium, Hungary, Portugal, Romania and Scotland) between 2001 and 2008. He argued that an instrumental use of the findings was widespread. PISA was employed ‘to justify reforms which were, to a large extent, already prepared or desired by providing political leaders with a new vocabulary for action’, even when these interventions ‘were sometimes contradictory and hardly based on the PISA survey’ (p. 215).

Gür, Çelik, & Özoğlu (2011) showed that, in Turkey, PISA 2003 results and related alarming headlines in the media were used by the Ministry of Education to push a curricular reform that had already been prepared. The authors also noted that some so-called ‘PISA-findings’ were common knowledge within the system.

In France, LSAs were traditionally viewed with scepticism, but PISA has progressively gained importance. Some authors stressed its usefulness in highlighting shortcomings in the system (Cytermann, 2010). Others reported that, since 2006, a ‘culture of international comparison’ and a ‘reform advocacy coalition’ developed (Dobbins & Martens, 2012, p. 32). Under the Ministry of Xavier Darcos, who had been the French ambassador at the OECD until 2007, PISA started to be used by the centre-right to advocate for performance-oriented reforms, and by the centre-left and the unions to direct the attention towards social inequalities (Dobbins & Martens, 2012).

The same happened in the French Community of Belgium, in which policymakers from opposite sides use the most convenient results to stress the inequality or the inefficiency of the education system (E. Mangez & Cattonar, 2009). Carvalho & Costa (2014) observe that policymakers tend to employ country statistics rather than policy advice from PISA in what they call a 'selective mobilization of certain "pieces" of PISA knowledge' (p. 4).

In all these cases, the OECD/PISA "brand" was used as a political stick, whereas OECD recommendations actually played a minor role in "PISA-based" reforms. The examples below contrast these findings with evidence that PISA can serve purposes other than instrumental accountability—though with varying degrees of success.

In Spain, PISA influenced some specific policies to promote reading and ICT in schools (Bonal & Tarabini, 2013). It also led to the creation of the "Ibero-American PISA Group", a 'mutual support group designed to contribute to a better performance on the PISA, by means of continuous technical training and exchanging best practices' (Bolívar, 2011, p. 67). Nevertheless, actual PISA-informed changes were not as common as the use of PISA to justify or obstruct reforms. This led Bolívar (2011) to comment that 'PISA results have not had any relevant pedagogical repercussions' (p. 65).

An interesting example is the Kyrgyz Republic. Here, 'the poor results of PISA 2006 gave the education officials an opportunity to strategically gain support from international development agencies' (Shamatov & Sainazarov, 2010, p. 158). Embracing the rhetoric of crisis, Kyrgyz officials 'learned to speak the language of the international donors and […] familiarized themselves with the current philosophy of aid that emphasizes needs rather than accomplishments' (*ibid*.).

In France, the political conflict mentioned earlier produced paradoxical results at times. A 2004 law aiming to reform the national examination by having part of the final grade depend on teacher judgement was informed by OECD recommendations. It was seen by the government as a way to make the assessment more reliable and fairer but it was rejected by students, who were afraid of teacher bias (Dobbins & Martens, 2012).

Japan and Korea have integrated 'PISA-type' tasks into their national examinations (Schleicher, 2009b, p. 492), but there is also evidence that in Japan PISA 2003 results were used to promote reforms whose rationale was actually in contrast with OECD recommendations. Takayama (2008, 2013) argued that while, on the one hand, the national assessment introduced in 2007 borrowed heavily from the PISA literacies, it disregarded the influence of socio-economic background on results and contrasted with the school-governance model recommended by the OECD. Similarly, the Danish government's response to the PISA 2000 unsatisfactory results 'emphasized a much more traditional conception of subjects […] than implied by PISA's disciplinary literacy […] framework' (Dolin & Krogh, 2010, p. 571),

and its decision to establish a new system of national assessment 'was not completely synchronised with the Danish PISA 2003 report' (p. 573).

At times, researchers disagree on the reception of PISA in the same education system. For instance, Berliner (2011) claims that 'people in the USA […] take test results from PISA very seriously' (p. 78), whereas others write: 'three cycles of the PISA study and their results have remained virtually unheard of' (Martens & Niemann, 2010, p. 15). This would be because PISA results do not depart enough from national expectations to prompt any reforms over and above what is already in the government's agenda (Bieber & Martens, 2011; Martens & Niemann, 2013). On this matter, Sellar & Lingard (2013a) have recently argued that the top performance of Shanghai in 2009 was the PISA-shock the US had been missing to begin the PISA-driven policy debate other countries had already experienced. The first outcome was an OECD report (Tucker, 2011) commissioned by the Secretary of Education 'on what the best performing school systems were doing and what lessons could be learnt from them' (Sellar & Lingard, 2013a, p. 474).

Neumann, Kiss & Fejes (2012) compared PISA reception in Hungary and Romania to argue that national policies were influenced not just by OECD recommendations, but by the particular attitude taken by national project managers with respect to PISA and its findings. During the development and administration stages of PISA, Hungarian project managers were committed and enthusiastic, whereas Romanian staff were disengaged and 'primarily conceived PISA as a formal bureaucratic task' (p. 233)—perhaps because country-participation was felt as a top-down imposition to 'align and comply with Europe' (*ibid.*).

The deeper internalisation of the PISA principles on Hungary's part translated into a rapid appropriation of the semantics and rhetoric of the OECD: PISA became a 'focusing event', used by the government 'to give external legitimacy' to controversial reforms, and by their critics to strengthen their positions (E. Neumann et al., 2012, p. 234; see also Kingdon, 1995, on focusing events). In Romania, in contrast, PISA was deemed incompatible with the system, so its contribution to systemic reform has been 'fragmented, slow and uncertain' (E. Neumann et al., 2012, p. 237).

The role of intermediate players such as PISA researchers and NPMs in producing knowledge for policy has also been recognised by Cattonar et al. (2009), Pons (2012) and Klemencic (2010). Klemencic (2010) reported that, in Slovenia, many professionals with responsibilities for implementing international LSAs at the local level also served as government advisors. Therefore, they had the ability to influence national education policies.

Pons (2012) argued that the PISA national centres 'played an ambivalent role' (p. 216), because their active participation in the production and dissemination of knowledge from PISA rarely resulted in '*knowledge for learning*' (p. 217, emphasis in the text). In other words,

intermediate actors were seen as collating and restructuring OECD evidence but without providing the tools to interpret it. This position contrasts with findings from the present case studies; it will be shown in Chapters 5 and 6 that the NPMs of Ireland and the French Community of Belgium actively provide the tools to interpret PISA reports.

To conclude this section, two in-depth summaries of the reactions to PISA in Germany and Finland are presented. Germany is widely cited as the country in which the influence of PISA was stronger, whereas the Finnish system is or has been a model to look up to; it is therefore instructive to see whether PISA has made any difference to Finnish policymaking.

The case of Germany

One of the main impacts of PISA 2000 is associated with the consequences it had in Germany—the so-called 'PISA shock'—though the country had experienced a 'TIMSS-shock' before (Kerstan, 2000; Seeber & Lehmann, 2013). Ertl (2006) argued that unexpected findings from PISA 2000 and 2003 about the inequalities of the German system and its underwhelming outcomes could be regarded 'as a watershed in the discourse on education in Germany' (p. 621).

The first effect was the rapid development and implementation of national educational standards, in line with the experiences of successful countries, and the establishment of the Institute for Quality Development in Education in 2004 to evaluate them. The standards drew heavily from PISA and represented a major shift in the German understanding of education, from a guided process of self-formation (*Humanistische Bildung*) to the skill-centred interpretation more common in English, American and Australian societies (K. Neumann, Fischer, & Kauertz, 2010).

Considering the German federal context, these changes 'would have been inconceivable without the impact of the PISA study' (Ertl, 2006, p. 623), but the policy impact on PISA was not limited to a single intervention: a 'post-PISA' agenda was published to address some of the shortcomings and inequalities highlighted in the international reports. Some scepticism was also expressed concerning the ability of the national standards to produce 'far-reaching improvement of teaching and learning' (p. 626), because they did not alter teacher dependence on existing traditional textbooks.

Tröhler (2011) argued that PISA 'created a market, for it has created customers with specific demands and specific supply' (p. 245). After the PISA-shock, consultancy and coaching actors such as the Bertelsmann Foundation initiated a series of educational projects to provide reassurance in a time of uncertainty (Hartong, 2012). Education stakeholders started to take research visits both to Finland and to Sweden (Ringarp & Rothland, 2010).

Sweden embodied an exemplar of autonomy, deregulation, equality of opportunities and effective pedagogy, and yet at the time of the German visits it was moving from a comprehensive to a selective system. It was 'in the process of dismantling just those parts of its educational policy that […] aroused interest and admiration in other countries, especially Germany' (Ringarp & Rothland, 2010, p. 422). This is an example of policy borrowing at a 'burnout' stage (Steiner-Khamsi, 2006, p. 666), that is, 'the borrowing of educational policies from other educational systems that, in the original context, were seen as failures, ineffective, or at least highly contested' (Steiner-Khamsi & Quist, 2000, p. 276; see also Halpin & Troyna, 1995).

The case of Finland

After analysing parliamentary documents since 1991, Alasuutari & Rasimus (2009) found that the OECD served four purposes depending on the occasion, so far as Finnish policymakers were concerned: it was a reference point to make between-country comparisons; a neutral body providing raw statistical data; a source of valuable and effective policy advice; and the voice of the international community and the 'ideological spirit of the time' (p. 99). Indeed, an analysis of Finnish educational reforms following five country reviews undertaken between 1982 and 2003, suggested that Finland had traditionally been very receptive to OECD advice (Rinne et al., 2004). The OECD provided the data for defining the ideologies and practices separating an in-group (OECD members) from the out-groups, and it set the parameters according to which the within-group competition could take place (Alasuutari & Rasimus, 2009). In fact, some referred to it as '"the instrument, catalyst and certain framework for comparison" for Finnish education policy' (Grek et al., 2009, p. 15; quoting Niukko, 2006).

It has been argued, however, that the influence of the OECD on Finnish policymaking may have been overemphasised. Rautalin & Alasuutari (2007) critiqued Rinne et al.'s (2004) contribution because it did not clarify whether Finnish reforms reflected an enactment of specific OECD recommendations or were simply aligned with some general goals and values but used internally-identified policy drivers.

Likewise, it is unclear whether the narrative about the excellence of the Finnish education system is reflective of reality. Simola & Rinne (2011) showed that three perceived truths about the factors behind the Finnish success in PISA—a high belief in schooling, the popularity of the teaching profession and the culture of trust towards teachers—are either supported by scarce evidence or are the product of contingencies rather than purposive planning. Sahlgren (2015) has recently added that the country's performance may have been due to elements pre-dating current education policies, such as socio-economic changes and traditional pedagogy.

With respect to the instrumental use of PISA data, moreover, Finland does not seem to behave differently from other countries. By examining the editorials of the main teacher union's publication in the 2001–2005 period, Rautalin & Alasuutari (2007) sought to understand how the 'exceptional' (p. 348) results in PISA 2000 and 2003 were used by the teaching community in a period of intense dissatisfaction with the system. They found that evidence from PISA was used to make claims regarding teacher professionalism in spite of inadequate resourcing. The union omitted the shortcomings in teaching practice that had emerged from PISA as they would have undermined the argument about the 'Finnish teachers' high level of professional expertise' (Rautalin & Alasuutari, 2007, p. 359).

Similarly, Rautalin & Alasuutari (2009) analyse how policymakers interpreted PISA results to justify existing or proposed interventions. As with the case of the teacher unions, the use of international outcomes was biased in the sense that the role of the government was overemphasised as a cause for success, whereas deficiencies were attributed to external factors.

Some considerations on the impact of PISA on national education policymaking

The increasing number of countries participating in PISA is no guarantee that the assessment is fulfilling the role it was conceived for: to help shape education reforms. The evidence reviewed in this section provides some clear pointers about the ability of PISA to drive policymaking, but it also opens some areas for further investigation.

PISA has certainly made an impact in the education policy discourse. References to PISA data are becoming increasingly common in political debates. Countries perceived as successful can achieve the status of 'reference' or 'world class' societies (Reynolds, Teddlie, & Creemers, 2002; Schriewer & Martinez, 2004) to which other countries can turn to seek an answer to their internal problems (D. Phillips & Ochs, 2004). However, one country could act as a reference society for another for historical reasons, and experts from within a reference society often question the degree to which the country is really successful. In fact, the role of reference societies is to support a narrative of crisis and reform, rather than to provide realistic solutions. They are 'invented communities' (Steiner-Khamsi, 2004b, p. 4); they can be any group that is 'out there' and is perceived to have certain qualities.

The society presumed to have these qualities can change over time (Carvalho & Costa, 2014; Waldow, Takayama, & Sung, 2014). This can be observed in PISA with the 'prodigal child' Finland being slowly replaced by higher-performing Singapore. At different times, different qualities are attributed to the same society. Takayama (2010) notes that, in Japan, the same education system (Finland) was used by parties to the left and the right to back their

respective arguments. Finnish education became ‘a multiaccentual signifier that is articulated into competing crisis-reform melodramas’ (p. 67). It is what Rappleye (2006) calls a ‘multivocal symbol’ (p. 233; reprising Turner, 1974) one ‘permitting different groups to “agree while actually holding different understandings and pursuing different agendas”’ (Goodman, 2003, p. 15; quoted in Rappleye, 2006, p. 234).

The OECD is claimed by policymakers to be objective and trustworthy, at least in public statements. To some extent, the international community represented by the OECD is itself a reference society, and policymakers often appeal to its authority to promote their viewpoints regardless of what the OECD actually recommended. Indeed, according to the literature, PISA is more often than not used as an accountability stick, a political instrument employed by governments and oppositions around the globe to back their lines of argument using only country scores and a few other indicators.

Steiner-Khamsi (2003) termed this approach to PISA “the politics of league tables”, whereby ‘international league tables […] are utilized as a policy tool to certify the demands of reform proponents’ (Steiner-Khamsi, 2004a, p. 208). Timing is a key component. Drawing from diffusion research (for instance, Meseguer & Gilardi, 2009), Steiner-Khamsi (2006) argued that reforms may spread like epidemics, and that if times are not ripe external shocks will be absorbed by the system. In her view, the meaning of international educational indicators ‘is determined domestically’; they are ‘“global speak,” instrumentally invoked at a particular time and in a particular policy context, to accelerate policy change’ (Steiner-Khamsi, 2010, p. 331).

In times of educational crisis, PISA rankings might be the external shocks that produce a window of opportunity for policy change (Breakspear, 2012; Kingdon, 1995). Martens & Niemann (2013) refine this framework and suggest that, to impact policy, PISA results must sizably deviate from the expectations at the national level and be framed as part of a wider problem (i.e., education crisis has to stand for economic or security crisis).

Note that PISA rankings are not used by reformists only; the ultimate purpose of an argument could be maintaining the current political course. Takayama (2008) reports that PISA 2003 ‘resonated with the specific cultural, political, and economic context of the time’, and that this was exploited by the Ministry of Education ‘to re-establish its political legitimacy in a time of increasing neo-liberal state-restructuring’ (p. 387). To policy defenders, ‘far from “dispelling myths”, PISA indicators may serve to legitimise preconceived notions of educational performance and problems’ (Henry et al., 2001, p. 96; see also Pons, 2012). Even Schleicher (2009a) concedes that ‘policymakers tend to use […international comparative benchmarks] selectively, often in support of existing policies rather than as instruments to challenge them and explore alternatives’ (p. 101).

It would be problematic for the OECD if the uses of PISA mentioned thus far were reflective of widespread practice, for two reasons. Firstly, it would imply that PISA does not return policy-valuable information by itself, it simply flags up national weaknesses and points to reference societies that supposedly do better. If this were the case, and PISA were simply a mirror of the state of education systems at one point in time, the role of the OECD would be reduced to that of being a data provider.

The second issue relates to the relevance of OECD recommendations. The PGB rejected a reduction of PISA to a ranking instrument and underlined on more than one occasion that its purpose is to give clear policy directions (OECD, 1999c, 2001b)— 'the *specific practices and policies* that provide more effective learning opportunities for students' (OECD, 2013h, p. 3, emphasis added). The OECD also acts upon critiques. For instance, after Bloem (2013) noted that PISA may be less useful for low- and middle-income countries, in 2013 the OECD launched *PISA for Development*—a project aiming 'to increase the policy relevance of PISA for developing countries through enhanced PISA survey instruments' (OECD, 2014j, p. 6).

Given that, why is it often the case that 'PISA results can be used to support A as well as the opposite of A [… whereby] he who holds the greater political power succeeds in using PISA as a justification for his own policies' (Figazzolo, 2009, p. 28)? Even the PGB was aware of the issue and following Figazzolo's (2009) report it decided to look into methods for preventing an instrumental use of PISA (OECD, 2009b).

To summarise, the issue at hand is that either OECD recommendations are not followed because PISA is, at best, a descriptive comparative survey returning ambiguous or overly-generic information, in which case the OECD's *ability* to provide accurate policy advice could be called into question; or they are ignored in spite of the fact that PISA really is a normative survey, in which case the *credibility* of the OECD as influential policy advisor could be doubted.

There is, in fact, a third case: some aspects of PISA may be descriptive and others normative. Some countries may instrumentally use of the assessment and others may have a more reflective approach. Cross-country evidence suggest that the use of PISA might not be as clear-cut and unsophisticated as sometimes has been claimed.

Popp et al. (2010) note that while PISA was accompanied by educational changes in Germany, France, Mexico, Switzerland and Spain, only in Germany did it cause major reforms; other countries showed only signs of 'adjustment' or 'alignment' (*Anpassung*, p. 302) to the perceived requests from the international community. Reactions also varied among English-speaking countries, from the lack of interest shown by the US (at least until now, Sellar & Lingard, 2013a) to renewed 'attention' or 'notice' (*Beachtung*, Popp et al., 2010, p.

302) in New Zealand and England (particularly after the establishment of the coalition government in 2010, see Baird et al., 2011).

De Olano (2010) suggests that there may be no single use of PISA results: in Japan, PISA results served to provide legitimacy to the government, whereas in Thailand they were used by critics to highlight its failures. PISA was integrated into the internationally-oriented reforms in Qatar, whereas a debate on its value is ongoing in Austria.

Drawing from a stakeholder's questionnaire, Hopkins, Pennock & Ritzen (2008) note that the influence of PISA on policy formation tended to increase from a country's second administration, though in the majority of cases the overall impact of the assessment was still medium or low. The reforms 'most likely to be adopted in light of PISA' (p. 35) included: changes in national curricular standards, the establishment of monitoring institutes, interventions targeting underperforming groups of students, changes in resource allocation and pilot projects. The authors identified three main policy actions following from PISA: curriculum development 'related to PISA domains and problem solving' (p. 75); the adoption 'of policy options from other countries and economies, but without systematic use of evidence' (*ibid.*); and the use of PISA as a stimulus to develop smaller-scale initiatives more sensitive to local contexts.

This led Hopkins et al. (2008) to conclude that 'although PISA undoubtedly has impact, the nature and form of this impact appears to be unpredictable' (p.74). Moreover, the perception of the impact level of PISA varied across stakeholders. While around a third of policymakers and local government officials believed that PISA had been 'extremely effective' in contributing to sustainable change by informing policymaking, less than 10% of academics and teacher unionists thought so, and even national project managers were somewhat sceptical.

This investigation was followed by Breakspear's (2012), who surveyed 37 members of the PGB. PISA was reported to have been 'extremely' or 'very' influential for policymaking in the majority of cases, regardless of actual country performance. The main effect concerned the 'formation, expansion and improvements' to national assessments, curricular reforms 'often to include and emphasise PISA-like competencies', and the introduction or amendment of performance targets (p. 15). Other interventions focused on socio-economically disadvantaged and low-achieving students. In many cases, policies were drawn from reference societies such as Finland, Korea, Shanghai, Singapore, Canada, Australia and New Zealand.

To address this section's opening question (How has PISA been received and used for national policymaking?), there is evidence that PISA is more than a simple ranking exercise and that it can be considered to a certain extent a public policy instrument (Lascoumes & Le Galès, 2007). At the same time, there is also evidence that PISA is not a "ready-to-use"

policymaking tool, either. Policy information from PISA seems to be either absent or routinely ignored—even though PISA presents itself as a product to be 'consumed' by policymakers (E. Mangez & Hilgers, 2012).

The double nature of PISA was summarised by von Bogdandy and Goldmann (2008). The authors set out from a legal perspective to argue that PISA can be framed as an exercise in public authority, because it exerts an impact on country policies that is sufficiently strong to limit the liberty of national actors. Nevertheless, they also note that actual policy recommendations were almost absent from PISA reports, and that is influence was 'an indirect one, generating general awareness that something must be done rather than inducing specific reforms' (p. 266).

Overall, the descriptive side of PISA appears to have both a use and an impact. Whether this is the case for its normative side, if any exist, is debatable. As mentioned earlier, this would put at risk the relevance of the OECD as a trustworthy policy advisor. In the remainder of this chapter it will be argued that, firstly, PISA does have a normative component: international reports and other PISA-informed OECD publications tell policymakers what they should be doing given PISA data. Secondly, there is a gap in the literature, specifically the fact that the contents of the normative side of PISA have always been assumed rather than explored.

PISA is a political exercise that aims to 'outflank' education systems 'by constructing parallel discourses and practices that at first sight do not challenge national sovereignty but nevertheless provide alternatives' (Dale & Robertson, 2007, p. 8). However, the literature has for the most part uncritically assumed that these discourses broadly dealt with deregulation, accountability and skills for life, without checking what the OECD actually *said* through and with PISA.

Instead, it is claimed below that the policy effects of PISA would be better understood if the scope of the investigation moved from 'what is the impact of PISA?' to 'what is the impact of PISA-informed education policy advice?'

The purpose, relevance and timeliness of this research

This research stems from the following observations.

A first observation is that although much research is dedicated to understanding the impact of the OECD operations in globalising national education policies through PISA, the contents of PISA-related publications have hardly ever been the starting point of the enquiry. What is the education policy advice (EPA) provided through PISA? What does the OECD suggest to its member countries that they should do?

It appears that part of the literature assumes the answer to be 'nothing' and the influence of PISA on country policymaking to be limited to the use and misuse of rankings. This view, however, minimizes the role of the OECD in the production of knowledge for policy and its interest in being at the centre of an epistemic community. PISA needs 'to meet the needs of a wide range of users – from governments wanting to learn policy lessons, through academics requiring data for further analysis, to the general public wanting to track how their nation's schools are progressing in producing world-class students' (OECD, 1999c, p. 3).

It is unclear from the literature what these policy lessons might be. All too often, OECD recommendations seem to have been dismissed as mere neoliberalism or neo-colonialism. This does not help in understanding what the OECD has to say about education. The political aspect of PISA is so prominent that some policymakers think it is all there is to it: 'PISA is an accountability engine. It tells you hardly anything about teaching and learning. It tells you that there is a problem but it doesn't tell you how to fix it' (interviewee in C. Morgan, 2011, p. 56).

From the perspective of the OECD and the Directorate for Education and Skills, this would be incredibly inefficient. Policy advice is the core business of the OECD: why not use a popular educational assessment to make educational recommendations? This author would argue that the OECD does use PISA-related publications for this purpose.

The PGB has been attentive to making PISA relevant and useful through an adequate dissemination strategy since the beginning, and many international reports have an "Implications for Policy" section. In fact, international reports are but one of the means at the OECD's disposal to reach its intended audience: PISA is the object of dedicated country and thematic reports, of 'a series of concise 4-page monthly education policy-oriented briefs [*PISA in Focus*] designed to describe a PISA topic', and PISA data are being integrated into other OECD publications, such as the EAG series. Moreover, PISA is referenced in online resources such as the *OECD Observer*, the *Education & Skills Today* blog and the new interactive database *Education GPS* (OECD, 2016a, 2016b, 2016c).

Even a cursory analysis of these publications reveals that the EPA exists. What remains to be analysed is what topics it covers and at whether the OECD policy recommendations may be valuable to education systems.

The EPA may be concerned with the quality, equity and efficiency of education systems—the core goals of the OECD. It might reflect the OECD strategy to disseminate a market view of education (Dale & Robertson, 2007), even though the PGB noted early on that PISA-informed reports should 'provide a perspective that goes beyond a labour market focus and exploits the full potential of the PISA instruments' (OECD, 1999b, p. 4). It might be 'tailored to the needs of OECD countries' (OECD, 1997a, p. 4), even though participation of countries

outside the OECD was always welcomed by the PGB (OECD, 1997a, 1998c). The fact is, with the notable exception of Bieber & Martens (2011) there have been no attempts to synthesise PISA-informed EPA in a coherent framework.

A second observation is that once a summary of the EPA is developed, research may proceed with questions such as: 'Are OECD recommendations consistent across time and national borders' (Armingeon, 2004)? 'To what extent are actual policy developments in a country in agreement with the OECD policy recommendations' (Beyeler, 2004, p. 1)? This would be the starting point for any serious attempt to move from the 'use and usefulness of PISA' (OECD, 2009b, p. 4) to *the use and usefulness of PISA-informed EPA*. It is a minor shift in focus that has major implications on research praxis and outcomes.

When the focus is on PISA, the enquiry concerns the construct validity and the technical qualities of the instrument or, as is the case with part of the literature referenced in this chapter, how rankings and information from the international reports were received at the national level and used as a justification for educational reforms. To some extent, this kind of research concentrates on the nature of PISA as a measuring instrument.

PISA, however, is also a public policy instrument, or at least is intended to be so by the OECD: its comparative aspect, which relates to how things currently are, subtends a normative aspect related to how things ought to be. On this interpretation, PISA is but a component in the OECD strategy to produce, disseminate and replicate educational knowledge. This shifts the focus onto the (implicit or explicit) policy advice accompanying the instrument.

Without any formalisation of the education policy advice, previous research framing policymaking under PISA had to borrow its interpretive lenses from other fields, hence concepts such as "neo-liberalism in education". It is argued here, instead, that the effectiveness of the OECD's attempt to steer education policies with the support of PISA is best explored by assessing the extent to which national policies match PISA-informed EPA.

A third observation, which is linked to the previous, is that when the focus shifts from PISA to PISA-informed EPA the concept of "usefulness" changes accordingly. "Usefulness" is defined in the Merriam-Webster Dictionary as 'the quality of having utility and especially practical worth or applicability' ("Usefulness," 2016), where the "utility" is in turn 'fitness for some purpose or worth to some end' ("Utility," 2016).

PISA may be useful for building national assessment capacity, connecting policymakers and making a large quantity of data available for research and secondary analyses. Wiseman (2010) argues that the usefulness of PISA is to focus public debate, to build capacity and to 'provide transparency in education where none may otherwise exist' (p. xviii). To Breakspear (2012), PISA is useful because it allows systematic evaluation in countries that do not have

national assessments in place, it provides an international benchmark that can be linked to national data, and it serves as a model for the design and development of national assessments.

To have fitness for some purpose or worth to some end, the EPA needs to possess different qualities from PISA. Policy advice from PISA must be reasonable, applicable and, most importantly, sound; it is useful to the extent to which following it aligns countries to the OECD goals by improving performance, equity and efficiency in education. The educational value of the EPA—that is, its worth as a reliable tool to inform effective education policies—is poorly understood because research has never followed the whole process of knowledge production, interpretation, use and impact from the 'international' to the national level in a systematic fashion.

Some work was carried out by the OECD, though the distinction between PISA and EPA was not made explicit. Late in 2001, the PGB saw in the evaluation of 'the impact and usefulness of PISA in Member countries an important medium-term perspective that would need to be considered from both policy and scientific perspectives' (OECD, 2002c, p. 8). A few years later, the PGB assigned to a group of experts the task of evaluating the use and the usefulness of PISA / EPA, and the outcomes were summarised in Hopkins et al. (2008).

Questions included the impact of PISA / EPA on country-level policymaking through 'a review of the delivery chain at national and international levels' (OECD, 2007d, p. 4) by assessing its influence on school-level practices and student learning, national assessments and policy coherence. The usefulness of PISA was to be evaluated by looking at its ability to provide added value to national assessment practices and 'a unique contribution for other forms of assessment' (p. 5). Finally, the evaluation was supposed to check 'the alignment of policies to the OECD strategic objectives over the longer term' (p. 4).

Hopkins et al. (2008) surveyed 548 interest groups ('policy makers, local government officials, school principals, parents, academics and researchers, and media representatives', p. 15) from 43 countries. They observed that 'one of the ironies of PISA is that although it focuses relentlessly on the achievement of students, it has relatively little to say about how the conditions of learning can be improved and less about giving school leaders tools to achieve this end' (Hopkins et al., 2008, p. 74) . Even at the central-policy level, there was a feeling that policy action has been unsystematic and incoherent, mainly because PISA results had provided little direction in this regard:

'The evaluation team […] strongly advises the OECD that it uses the outcomes of PISA to stimulate a more precise debate among participating countries and economies on how various policy options relate to outcomes on PISA assessments. […We] would suggest that *as a minimum, PISA considers the creation of a policy group for countries who request advice*

on policy formation in light of PISA results' (Hopkins et al., 2008, pp. 74–75, emphasis in the text).

Measuring policy impact, and particularly the impact 'of soft governance, is challenging […, and] the question remains to what extent countries would have implemented the same measures irrespective of the OECD's governance mechanism' (Martens & Jakobi, 2010, p. 272). Moreover, 'it is difficult to decipher the direct impact of PISA on the speed or breadth of reforms enacted' (Breakspear, 2012, p. 6). As a result, the Hopkins et al. (2008) evaluation commissioned by the OECD remained an isolated attempt.

The present evaluation aims to analyse and assess the nature, use and value of the EPA by relating a complex knowledge-production process (PISA-informed EPA), to its chain of implications at the national level (the development and implementation of EPA-informed policies), then to the endpoint: the effectiveness of national policies in raising student performance and equity.

More specifically, drawing from the three observations in this section, this research seeks to:

1) (Re)construct the OECD education policy advice, analyse its internal consistency and soundness, and explore the extent to which it is actually informed by results from PISA.
2) Analyse the degree to which national education policies are aligned with PISA-informed EPA, rather than merely exploit the OECD "brand" or PISA rankings to justify their existence. This would address a request for 'a more thorough study of the temporal relationship between policy thrusts and PISA results […to] shed light on the plausibility of the claim that PISA causes policy' (Baird et al., 2011, p. 2, emphasis added).
3) Begin to understand the plausible impact of EPA-informed policies on student performance or the equity of the system, because 'whether and how PISA has helped develop education (systems) and its consequences for school practice merit further research' (Hanberger, 2014, p. 176).

The rationale for point 3, which is schematised in Figure 1.2, is:

If

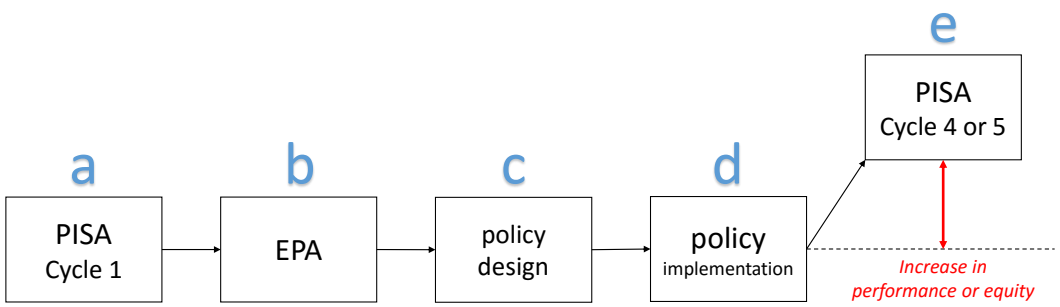
a. PISA helps to identify certain relationships between systemic features and student outcomes, and
b. The EPA draws from PISA findings, research and best practice to suggest ways to address shortcomings at the country level, and
c. national policies draw from the EPA, and

d. they are successfully implemented;

Then

e. After some time, PISA should be able to detect some improvements in the system.

Figure 1.2: How PISA is meant to help education systems improve.



Source: own elaboration.

A critical observation must be made: the model in Figure 1.2 is an ideal. *There is no doubt* that the environment in which PISA operates is so complex that policy generation and implementation cannot proceed in such a linear fashion. Nevertheless, it is argued here that the closer the actual process to this model, the greater the value of PISA-informed education policy advice and the usefulness of PISA. Likewise, it is argued that deviations from this model can be—if not measured—at least empirically observed, categorised and perhaps ordered, and that the results of such exercise could be used for increasing the utility of OECD recommendations.

The next chapter introduces the guiding hypotheses, theoretical framework and design of this research, as well as its leading questions.

Chapter 2: Research methodology

This chapter is articulated in the following parts:

- A rationale for the research and its objectives, drawing from the previous chapter.
- The theory framing the research, including a working definition of "education policy advice" and an overview of the rhetorical devices employed by the EPA, to guide the analysis of the OECD discourse.
- The research questions and hypotheses.
- A detailed presentation of the mixed-methods research design, broken down by research question (a schematic overview is also available in Appendix 1).
- Ethics considerations concerning the use of expert interviews as a method of data collection.

Rationale and objectives

In the previous chapter, it was argued that PISA has thus far played a double role: an "accountability stick" in the politics of league tables, whereby PISA rankings and indicators are selectively used to back contrasting policy positions; and a public policy instrument, making recommendations which may influence country policies. It was noted that countries tend to employ PISA more often in its first role, whereas the ambition of the OECD seem to be that the normative aspects of PISA prevailed in the policy discourse.

Since it is unclear whether the effect of OECD recommendations is low or simply under-researched, a shift in the scope of the investigation from the use and usefulness of PISA to the use and usefulness of the PISA-based education policy advice (EPA) was advocated. The few attempts to reconstruct the EPA came from political sciences and sociology—meaning that educational recommendations were taken as a "proxy" for the OECD's neoliberalism or globalising effects rather than as something worth exploring in itself.

The chapter concluded by suggesting that framing the EPA as it emerges from OECD publications (and PISA reports in particular) could be the starting point in assessing the educational value of OECD recommendations: the degree to which PISA-informed policy advice manages to translate into clear policy recommendations that can be (and have been) reliably used at the national level to inform effective educational interventions.

This research could be framed as an exercise in "education policy advice evaluation". Its purpose is to suggest ways of improving the validity of both PISA and the EPA by linking them to consequences and outcomes, while providing the theoretical and methodological

instruments to extend this work to other countries and future assessments. It aims to contribute to the fields of policy evaluation and (international) assessment validation.

Theoretical framework

As an evaluation of education policy advice, this research is placed at the intersection between the assessment validation and the policy evaluation fields. From validation theory, the research applies the concepts of "validity" and "reliability" to PISA; from evaluation theory, the research applies the concept of "programme theory" (Chen, 1990) to the EPA. It also agrees with realistic evaluation (Pawson & Tilley, 1997) that no policy 'just works': policy effectiveness has to take into consideration the context of implementation to interpret observed outcomes. The following section expands on the contribution of both fields and clarifies how they are linked by and within this study.

Validity and validation

The value of the EPA is linked to the validity of PISA, because the strength of the OECD recommendations relies on what can be plausibly inferred from PISA. If PISA indicators were inaccurate or unreliable over time, one could not claim that some features of education systems produce certain educational benefits.

The concept of validity has progressed in the last century from being an inherent property of a test—whereby a test is valid if it measures what it purports to measure (Kelley, 1927)—to being a property of the theory behind a test (Cronbach & Meehl, 1955), to being a property of the inferences that can reasonably be drawn from a test and actions taken as a result of the test: 'Validity is an integrated evaluative judgement of the degree to which empirical evidence and theoretical rationales support the *adequacy* and *appropriateness* of *inferences* and *actions* based on test scores or other modes of assessment' (Messick, 1989, p. 13, emphasis in the text).

The phrasing is not ideal, because Messick used the word "validity" to indicate both a process (the 'integrated evaluative judgement'—i.e., *validation*) and its outcome ('the degree to which…'—*validity*). Nonetheless, the theorisation is clear: validation is about assessing whether test interpretations and uses are warranted by what a test is designed to measure and how it purports to measure it. Under this definition, the test properties and technical features become part of the evidence in support of its validity; this includes its reliability, which is the capability of a test to produce similar sets of scores given similar inputs.

Messick's definition of validity was not uncontested. The debate mainly focused on whether social consequences such as 'unanticipated side-effects of legitimate test use' (Messick, 1998, p. 40, emphasis removed) should or should not be part of the validation

process. Proponents of a "socio-politically neutral" nature of validity tried to separate validity from validation—for instance by arguing that validity is a property of the test, and that a test is valid 'if and only if […] the attribute exists and […] variations in the attribute causally produce variations in the outcomes' (Borsboom, Mellenbergh, & van Heerden, 2004, p. 1061).

The counter-argument to the socio-politically neutral stance was that narrowing the scope of validity would limit its usefulness: 'It is rare that anyone measures for the sheer delight one experiences from the act itself. Instead, all measurement is, in essence, something you do so that you can use the outcomes…' (Hubley & Zumbo, 2011, p. 219; quoting Zumbo, 2009, p. 66). The intended uses of LSAs are often made explicit by the developing organisations. IEA studies have traditionally aimed 'to use psychometric techniques to compare outcomes of different education systems' (Mullis & Martin, 2006, p. 2), to provide data to identify the 'key factors that influence teaching and learning', but also to 'direct educational reforms' and to develop 'educational systems' capacity to engage in national strategies for monitoring and improving performance' (IEA, 2010). PISA, instead, can be used to compare outcomes and targets, analyse trends and support 'the political economy of reform' (Schleicher, 2009a, p. 103).

The debate on whether social consequences should pertain to validity is ongoing and has recently been reviewed by Newton & Shaw (2014), but this research is more concerned with the 'inferences' rather than the 'actions' (Messick, 1989) from the survey. For the most part, the validity of PISA is critiqued in Chapter 4. Some limitations concerning the assessment design and its reliability are identified and used to argue that many of the country score fluctuations might be a statistical artefact rather than capture policy effects, as is instead often argued.

The link between assessment validation and policy evaluation

One strand of the research on validity provides a connection between test validation and policy evaluation. It is this strand which allows a shift from the validity of PISA to the validity of the EPA, which is tackled in Chapter 3.

Kane (2001, 2006, 2013a), drawing, among others, from Toulmin's (1958) model of argumentation, developed an argument-based approach for test validation, whereby the proposed interpretations and uses of a test should be made explicit and summarised in an 'interpretation/use argument' (M. T. Kane, 2013a, p. 2) which then serves 'as a template for validation' (M. T. Kane, 2013b, p. 17). Explicit interpretation/use arguments are already available for PISA, because international reports provide both an extensive analysis of country outcomes (interpretation) and some policy recommendations (use). Therefore, it is theoretically possible to validate the EPA using Kane's (2001, 2013a) approach.

Consider that the EPA, which is the "end product" of PISA (and as such can be investigated through the lens of validation theory), is also intended to be one of the sources of national education policies, because countries are supposed to act upon OECD recommendations. This is where the literature moves from test validation to policy evaluation.

A key step in theory-based evaluation (Fitz-Gibbon & Morris, 1996) is the analysis of the programme theory (Chen, 1990) or programme logic (Funnell, 1997) of the intervention. The programme theory represents the set of assumptions and mechanisms through which the policy or intervention is expected to cause the intended or observed outcomes, and it is the model guiding the evaluation (Rogers, Petrosino, Huebner, & Hacsi, 2000; Rossi, Freeman, & Lipsey, 2004).

Since the EPA *is* a set of mechanisms through which countries are supposed to improve their performance, a connection between argument-based test validation and theory-driven programme evaluation emerges: not only the two fields share one object of investigation (the education policy advice, first link), but they also employ similar strategies to analyse it (second link). Overlaps can be found even in the language of the scholarship: for instance, in assessment validation, Kane (2013a) argues that the inferences about the interpretations and use of scores 'take the general form of "if-then" rules' (p. 11); likewise, the 'policy-scientific approach' of evaluation (Leeuw, 2003) requires that statements about causal mechanisms should be reformulated 'in conditional "if–then" propositions or propositions of a similar structure' (p. 7).

These similarities are not surprising if one considers that the EPA is a 'speech act' (Austin, 1962)—or, more precisely, a (directive) illocutionary act—that is, an attempt 'by the speaker to get the hearer to do something' (Searle, 1976, p. 11). OECD recommendations are both an argument (whereby implementing certain policies will bring benefits) and an intervention (i.e., an attempt to intervene in national policy formation): their argumentative side can be validated, whereas their illocutionary effectiveness can be evaluated.

The use of policy evaluation methodologies to validate (inferences from) international assessments has a precedent: for instance, Hanberger (2014) evaluated the programme theory behind the PISA enterprise. He assessed whether the assumptions behind the ability of PISA to influence the policies of participating countries were supported by the literature. He found that 'there is no logical connection between […] the activities of publishing PISA results […] and providing policy recommendations […] and the assumption that these activities will help actors, particularly at the local level, learn about their own practices and take action to improve education' (Hanberger, 2014, p. 177). Therefore, he concluded that 'PISA can at best function as an alarm system and a facilitator of policy change at the national level' (*ibid.*).

Notice that Hanberger (2014) connected assessment validation and policy evaluation by analysing the use of PISA 'as a programme for transnational governance and an evaluation system producing evaluative knowledge for changing national education policy in line with the OECD's objectives' (p. 170); this research, instead, focuses on the use of the EPA as the product of the interpretation of PISA findings and a tool to improve the quality of education in participating countries.

The difference is subtle but can be expressed as follows: while Hanberger (2014) theoretically evaluated whether PISA-informed EPA *could influence* national policies, this research adds one step and empirically evaluates whether PISA-informed EPA *has influenced* national policies, and whether EPA-informed policies *could influence* country performance.

Developing a theoretical framework for the education policy advice

As mentioned in these first two chapters, so far there has not been any systematic attempt to frame the *education* policy advice of the OECD. For this research, it was necessary to draw from a range of fields to define what education-specific policy advice is and what it may look like in the OECD discourse. The "ontology" of the EPA was primarily informed by evaluation literature, whereas its "semiotics" by political discourse theory.

The nature of education policy advice

There is disagreement about what exactly qualifies as a "policy", but there is some consensus that policies are both a textual product and an unwritten process; they are systematic stances or sets of practices that governments and other stakeholders adopt to address problems, (re)distribute resources or (de)regulate behaviours (Taylor, Rizvi, Lingard, & Henry, 1997).

Policymaking is not limited to parliamentary (primary) legislation, but includes secondary legislation such as ministerial orders, regulations and all official decisions made by a delegated authority (for example, a governmental agency). Policies can be more or less prescriptive. They can be very wide in scope and time, such as national plans, frameworks or fundamental acts, or be limited to short-time programmes or interventions targeting small groups or areas.

Since this research aims to uncover the specific education policy advice of the OECD, one would think that adopting the OECD definition of "policy" would solve the terminological issue, but there is no such definition, even though the organisation writes as if the meaning was clear. The OECD seems to have accepted Sir Charles Cunningham's adage whereby 'policy is rather like the elephant - you recognize it when you see it but cannot easily define it' (Cunningham, 1963, p. 229).

This research adopts instead the definition of Knoepfel et al. (2007), whereby policy is:

> 'a series of intentionally coherent decisions or activities […] carried out by different public […] actors whose resources, institutional links and interests vary,

> with a view to resolving in a targeted manner a problem that is politically defined as collective in nature. This group of decisions and activities gives rise to formalised actions […] aimed at modifying the behaviour of social groups presumed to be at the root of, or able to solve, the collective problem to be resolved (target groups [for instance, the teachers]) in the interest of the social groups who suffer the negative effects of the problem in question (final beneficiaries [for instance, the students])'. (Knoepfel et al., 2007, p. 24, emphasis removed)

The 'formalised actions' Knoepfel et al. (2007) refer to are related to the concept of "policy lever". Policy levers are the compulsory (e.g.: legislation, regulation, official texts) and non-compulsory instruments (resource allocation, persuasion or pressure) that policymakers can use to translate their policies (in the more abstract sense of stances, wants and sought practices) into action. They are the expression of a political will. Levin (2005) categorised policy levers as in Table 2.1.

Table 2.1: Levin's categorisation of policy levers

Type	Example	Means
Mandate	Primary and secondary legislation	Legal coercion
Inducement	Additional funding, recognition of accomplishment	Strategies to promote attention to policy goals
Capacity-building	Training	Building of institutional skills and systems
System-changing	Reform of bodies or agencies	Changes in structures
Opinion mobilization	Communication strategies	Non-mandatory external pressure

Source: own elaboration from Levin (2005, pp. 152–153)

There is a certain consistency between the OECD's and Levin's (2005) use of the phrase "policy levers". Both seem to employ it to indicate political actions and choices. For instance, EAG distinguishes between 'learning outcomes […], policy levers or circumstances that shape these outcomes, [… and] antecedents or constraints that put policy choices into context' (OECD, 2014d, p. 17). The OECD adds: 'antecedents at a lower level of the system may well be policy levers at a higher level. For teachers and students in a school, for example, teacher qualifications are a given constraint while, at the level of the education system, professional development of teachers is a key policy lever' (p. 18).

The ability of Levin's (2005) framework to capture the OECD's definition of "policy lever" was tested using a document explicitly designed 'for policy makers, analysts and practitioners'— the Education Policy Outlook for Australia (OECD, 2013b, p. 2, emphasis removed)—by checking the extent to which Australian 'policy responses' (p. 4) identified by the OECD fell within Levin's (2005) model. Upon analysis, most policy responses (mainly national strategies, partnerships, frameworks, standards and funding allocations, but also small-scale programmes) belonged to one of Levin's (2005) categories. The most difficult interventions to frame were government-commissioned reviews of current practices: in fact,

these can be considered precursors of policies aimed at drawing attention to an issue and putting accountable stakeholders under pressure, rather than policies in their own right.

Once an operative definition of "policy" was linked to the contributions of Knoepfel et al. (2007) and Levin (2005), the scope moved onto the "education" part of the phrase. Intuitively, education policies are those that fall within the remit of a Ministry of Education [MoE][7], but many "education" policies nowadays are the result of collaboration among different offices, units, departments or even ministries.

Typical examples are large social protection policies meant to help children in education, such as the *Child Poverty Strategy* in the UK (Department for Work and Pensions & Department for Education, 2011) or the smaller-scale *Back to School Clothing and Footwear Allowance* in Ireland (Department of Social Protection, 2015), which 'helps […parents] meet the cost of uniforms and footwear for children going to school' but is dispensed by the Department of Social Protection. Other examples include instances in which a Ministry, like the Ministry of Health, draws from the expertise of the MoE to deliver education on health-related issues (this was the case in Saudi Arabia; see MOH Portal Team, 2014); or conversely, when the MoE asks more competent Ministries to help with the definition and implementation of specific policies (like an IT policy in Moldova; Moldovan Association of ICT Companies, 2014).

For its part, the OECD has advocated more policy synergy since the 1990s and is an active promoter of policy coherence for sustainable development, whereas PISA regularly measures background variables, such as age of arrival in the country or number of TV sets available in the house, which are hardly "educational" in the strict sense but can capture family contexts and mechanisms (such as parental ability to negotiate their children's educational needs at home) that affect academic performance. In summary, if one were to judge from real-life situations, it would not be evident whether 'education' in "education policy" should be taken to mean policies for the education system, for people in education or for the education of people.

For practical reasons, the scope of this investigation was limited to policies for which the MoE could be held accountable. Furthermore, to limit instances of social and welfare policies stretching this boundary, this research focused only on policies involving schools and school staff directly, or seeking to educate parents with a view of influencing their children's behaviour or performance in school.

Thus, for the purpose of this document, "education policy" is defined as:

[7] In many Anglo-Saxon countries the equivalent to the Ministry of Education (MoE) is the Department of / for Education.

> ‘A decision or activity as intended by Knoepfel et al. (2007) giving rise to some action that can be broadly captured by Levin’s (2005) policy lever framework and for which the Ministry of Education (or a delegated agency) is primarily responsible.’

By extension, “education policy advice” is generic or country-targeted recommendations involving education policies the OECD makes[8].

The form of education policy advice

The second step for capturing the EPA in the OECD discourse was to understand the form it might take in terms of rhetoric style and textual features. The theoretical approach chosen drew from the concept of “practical arguments” (Fairclough & Fairclough, 2012; a link with assessment validation is provided in M. T. Kane, 1992; originally in Toulmin, Rieke, & Janik, 1979). It was hypothesised that OECD recommendations would share at least three qualities with practical arguments.

Firstly, practical arguments are more concerned with action than with truth. The conclusion of a practical argument is not about ‘what is’ but rather about ‘what should be done’. Because of this, in ‘political and other practical argumentation’, even when there is not ‘anything wrong or questionable about an argument’ the other party ‘may be entitled to reject the proposal for which the argument is presented in support’ (Kock, 2007, p. 98)[9]. Indeed, the effectiveness of a practical argument does not rely so much on truthfulness and falsehood as it does on the ability of the argument to appeal to the hearers’ goals and values.

Secondly, unlike classical argumentation such as induction or deduction, practical arguments are based on ‘plausible’ and ‘presumptive’ inferences: a plausible inference ‘is one that can be drawn from the given apparent facts in a case suggesting a particular conclusion that *seems to be true*’ (Walton, 2001, p. 166, emphasis added); a presumptive inference ‘enables a conclusion to be drawn *provisionally* from premises’ (*ibid*., emphasis added), provided ‘there is no sufficient evidence to show that the proposition is false’ (Fairclough & Fairclough, 2012, p. 39; quoting Walton, 2006a, p. 72). It will be shown in the next chapter that much of the EPA is based on presumptive inferences.

Finally, both OECD recommendations and practical arguments are used in “deliberations”, ‘a group activity in which it is assumed that although the participants have differences of opinion, they also share common goals, and want to move ahead in taking a

[8] As an aside, note that even some actions of the OECD, such as peer reviews, could fit into Levin’s (2005) framework—specifically in the last category. This suggests that OECD governance methods may count as education policy, blurring even more the boundaries between “international” and “national” policymaking.

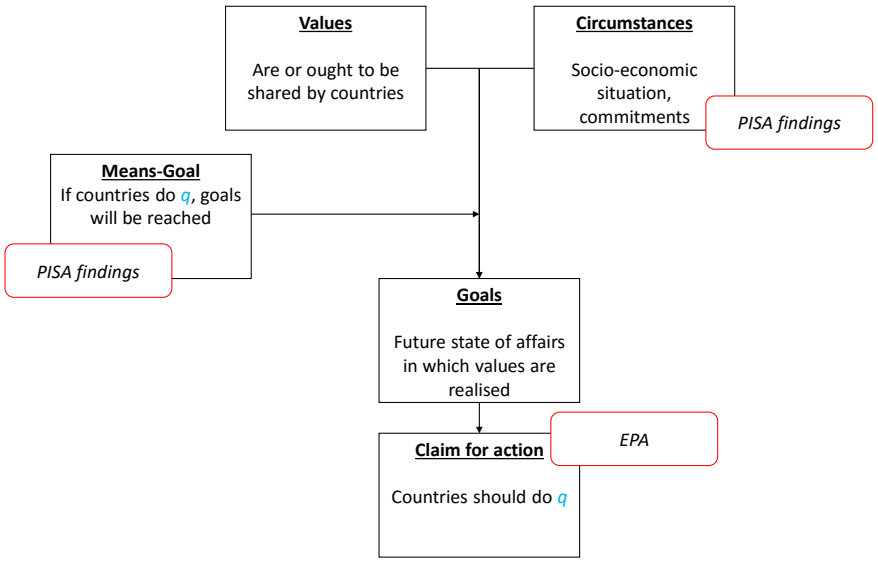
[9] There is a technical difference between a “proposition” and a “proposal”: since proposals concern what to do, their conclusions cannot be true or false (Kock, 2007; also Russell, 1951). In this section the two terms are used interchangeably.

collective action to carry out some task' (Walton, 2006b, p. 181). This recalls the consensus-building activity of the OECD. The EPA is an attempt 'by the speaker to get the hearer to do something' (Searle, 1976, p. 11), but it takes place within a deliberating context, as there is an understanding that all hearers should follow the advice for the common good.

The argumentative structure of deliberations was called by Walton (2006b) 'multi-agent practical reasoning' and was summarised as follows: 'We have a goal *G*. […] *G* is supported by our common set of values, *V*. […] Bringing about *q* is necessary (or sufficient) for us to bring about *G*' (p. 214). Fairclough & Fairclough (2012) adapted this model to political discourse with two modifications: they included the limiting or enabling circumstances in which decisions have to be made, and they recognised that sometimes the sharing of goals and values may be requested rather than pre-existing (it is not a matter of 'I / we want' but of 'you ought to want' something for the common good).

These contributions informed a preliminary model of multi-agent practical reasoning applied to the OECD (Figure 2.1). PISA results feature twice in the picture because of their double role: on the one hand, they monitor the current situation (the circumstances) of each country; on the other hand, they are meant to provide evidence that certain behaviours are associated with certain outcomes, with an implication that adopting the behaviour would make it easier to reach the goals. Walton (2006b) calls this component the 'Means Statement' whereas Fairclough & Fairclough (2012) the 'Means–Goal' (i.e., how the means are linked to the goals).

Figure 2.1: Draft model of multi-agent practical reasoning applied to the OECD



Source: Own elaboration informed by Fairclough & Fairclough (2012); Walton (2006a)

The ability of the draft theoretical model in Figure 2.1 to capture instances of EPA was tested with a practical argument extracted from *PISA in Focus*. The analysis is presented in Appendix 3 (Table 2.A) and the exceept is the following:

> 'PISA results consistently show that a 15-year-old student – regardless of his or her background – who had attended pre-primary school for at least one year performs better in mathematics than a student who hadn't. That pre-primary enrolment rates are growing faster among advantaged students than among disadvantaged students signals that countries have to work harder to ensure that all families, particularly disadvantaged families, have access to high-quality pre-primary education, and to information about such programmes, near where they live. An investment in early education, both for parents and for governments, pays dividends later on in life.' (OECD, 2014c, p. 4)

The analysis of this paragraph highlights some shortcomings in the ability of the preliminary model to frame the EPA while revealing many salient features about the way the OECD delivers policy recommendations.

Firstly, the EPA is rather specific even though it is directed to a very wide audience. What countries should do is explicitly mentioned: target pre-primary education, both in terms of quality and quantity, through investment and information. The strategy that each individual country would adopt to target pre-primary education through investment and information is left open, but it is likely to fall within one of Levin's (2005) categories (Table 2.1). This confirms that the excerpt above has the intent of informing education policies, and therefore that this is a legitimate EPA instance.

Secondly, there are some core underlying goals and values that provide the basis for the whole argument (performance, equity, efficiency) but, unlike the EPA, they are often left implicit, as they are probably assumed to be shared by the recipients. Evidence from PISA is used to support both the Means–Goal and the Circumstances parts of the argument. If countries ensure that families have access to pre-primary education (Means-Goal) the goal of having higher performance (implicit in the text) will be reached because students 'who had attended pre-primary school for at least one year performs better in mathematics than a student who hadn't' (PISA findings). Likewise, the fact that 'pre-primary enrolment rates are growing faster among advantaged students' (Circumstance and PISA findings) suggests that countries should 'work harder' to increase access to pre-primary education for disadvantage families (Means-Goal) to reach the valued goal of equity (implicit in the text).

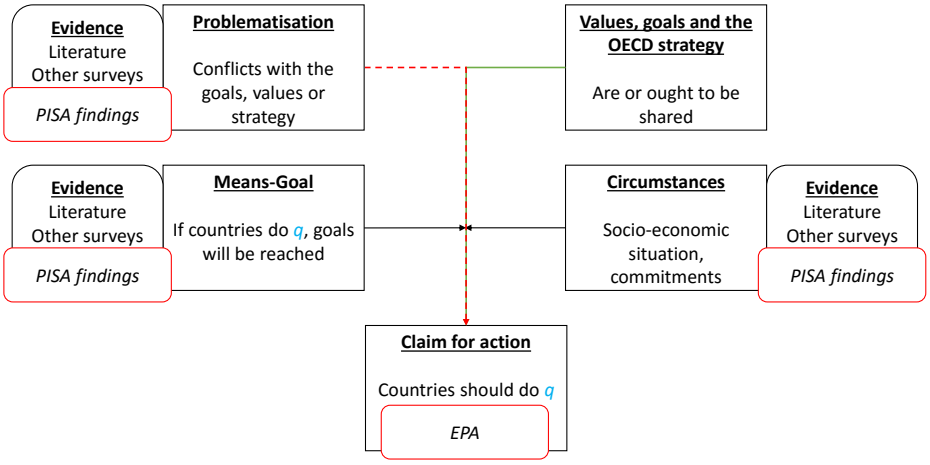
Notice, however, that not all recommendations follow from PISA evidence. For instance, the fact that pre-primary education has to be 'high-quality' is a reasonable suggestion but it is not and cannot be based on PISA results, because quality is not a variable measured by the survey (whereas length of enrolment is). This suggests that, even when commenting on

findings from PISA, the OECD may interweave evidence from PISA with that from other sources (including, but not only, the educational literature).

One limitation of the model above was that it did not capture satisfactorily a specific function of PISA findings that was frequently highlighted in the literature: signalling that there is an issue. When the OECD mentions that growth in enrolment is higher among advantaged students, it is not only describing a state of affairs (which is what Fairclough & Fairclough, 2012 mean by "Circumstances"); it is implying that the state of affairs is unwelcome. Indeed, the sentence 'pre-primary enrolment rates are growing faster among advantaged students' accomplishes two purposes: it *problematises* a situation and at the same time it *delimits* (circumscribes) the relevance of the argument to policymakers who feel more affected by the problem. Countries in which access to pre-primary education is felt (with or without reasons) as a current and real social issue will be more likely to be receptive take heed of the advice.

As a result of these considerations, the draft model in Figure 2.1 was revised to emphasise the problem-framing aspect of PISA.

Figure 2.2: A revised model of multi-agent practical reasoning applied to the OECD



Source: Own elaboration informed by Fairclough & Fairclough (2012); Walton (2006a)

The revised model considers the EPA as the endpoint of an argument beginning with a tension between the aims of the OECD and some problematic factors that prevent achieving the goals and realising the values. The EPA is the proposed solution for the clash. PISA is but one of the pieces of evidence used to support parts of the argument. Data from PISA can be used to flag up and delimit the problem or to sustain the means-goal statement with the experience of high-achieving countries.

The appropriateness of this model was confirmed by its superior ability to frame the previous excerpt from *PISA in Focus* (Table 2.2, below; a further analysis is presented in Appendix 3, Table 2.B).

Table 2.2: The revised practical reasoning model applied to an OECD argument from PISA in Focus *n. 40*

Problem	**Evidence**
'Pre-primary enrolment rates are growing faster among advantaged students than among disadvantaged students'	PISA
Goals and values	
Performance, equity and efficiency in the education system	N/A
Circumstances	
Same as problem	N/A
Means-goal	
If (low-SES) families can access pre-primary education (equity and) performance will improve	'PISA results consistently show that a 15-year-old student – regardless of his or her background – who had attended pre-primary school for at least one year performs better in mathematics than a student who hadn't'
If quality is high, performance will improve, because quality improves performance	N/A
If low-SES families know about pre-primary programmes equity and performance will improve, because more information will lead to more access and therefore to higher equity and performance	N/A
'An investment in early education […] pays dividends later on in life'	N/A
EPA	
'Countries have to work harder to ensure that all families, particularly disadvantaged families, have access to high-quality pre-primary education, and to information about such programmes, near where they live'	N/A

Source: (2014c, p. 4). The parts in *italic* are implicit and can only be presumed to correspond to the OECD line or reasoning.

This preliminary analysis gives a sense of what to expect when looking for PISA-informed EPA in international reports and other publications. OECD recommendations are explicit but values and objectives are often implicit. In line with findings from previous literature, PISA is used to define problems and provide solutions. But the most interesting feature in the OECD model of argumentation is that, as it happens in the political arena, not all claims are backed by PISA evidence, and the transition between supported and unsupported claims happens frequently and tacitly.

Of course, while the purpose of PISA is to provide 'evidence-based policy guidance' (OECD, 2011a, p. 5), there is a limit to the number of variables it can measure. But if even the OECD, whose core values include being objective and evidence-based (OECD, 2011a), does

not clearly separate what was measured from what was assumed, how can it criticise policymakers (e.g., in Schleicher, 2009a) for using PISA instrumentally?

Research questions and a working hypothesis

This research is an exercise in education policy advice evaluation. It draws from evaluation theory to define "education policy" and adopts the conceptual instruments of political discourse theory to identify the EPA. The EPA is considered to be the source programme theory for national education policies, and in agreement with realistic evaluation, national policies are investigated in context. The validity and reliability of PISA are also explored as part of the evaluation.

The research is guided by four questions:

RQ1) Is the OECD education policy advice "sound" advice?

RQ2) To what extent do countries follow the OECD education policy advice in their national policies?

RQ3) What is the evidence that education policies informed by OECD recommendations influenced country performance?

RQ4) Considering the evidence, how valuable is the OECD education policy advice?

RQ1 aims to reconstruct the EPA and analyse whether it is coherent, consistent and supported by evidence (including PISA evidence). This is the more theoretical aspect of the OECD policy advice. Its practical side is investigated in RQ2 and RQ3. RQ2 concerns the use of the EPA. Acknowledgement of the OECD goals and values by policymakers is separated from policy alignment with the EPA. Only reforms that are actually in line with the EPA can provide evidence that following it makes a difference for country performance and equity. RQ3 searches for this evidence. Finally, the last research question summarises the evaluation by looking at the overall value of the EPA. Here, "value" means that the EPA is advice worth following either because there is evidence that it *could* work (RQ1) or because there is evidence that it *did* work (RQ2 and 3).

One possible set of answers to these questions are summarised in the following box, which can be used to make explicit the researcher's stance, as well as his bias, during the investigation:

Education policy advice might be vague or inconsistent over time. Despite the powerful narrative about the influence of the OECD in national policymaking, reforms that seem aligned with OECD recommendations may not necessarily informed by them. Of the few policies that make explicit use of the EPA, even fewer might be successfully implemented.

Finally, a very small number of implemented policies may have plausibly contributed to (positive or negative) changes in educational outcomes, even when these outcomes are measured by the very assessment that originally informed policymaking.

Whether the argument above holds or not depends on the findings from this research. It is just one hypothetical line of reasoning that needs being proven. To summarise it, is it possible that:

Despite the use of data from PISA within political and scientific circles, OECD recommendations are currently of little educational value, because they are not solidly grounded empirically; they provide vague policy direction; advice is not followed; and even when it is followed it is not associated with measurable changes?

Research design

The main components of theory-based evaluation are the explication of the programme theory and of the causal links between the theory and the outcomes, as well as the gathering of theoretical and empirical evidence to 'assess the strengths and weaknesses of the links in the theory' (Mayne, 2012, p. 272). The objective is 'not only knowing whether a program is effective […] but also explaining a program's underlying causal mechanisms' (Coryn, Noakes, Westine, & Schröter, 2011, p. 203).

These components are integrated into this research as follows. Firstly (RQ1), the programme theory is made explicit by reconstructing the EPA, which provides the overall argument linking country policies to the shared goals of performance and equity (measured in terms of PISA scores). This part of the research is covered in Chapter 3. Secondly, evidence is sought of country alignment with OECD recommendations (RQ2) and of the effectiveness of EPA-aligned policies (RQ3). Both questions are tackled in Chapters 4–6. Chapter 4 addresses the issue from a cross-national perspective, whereas Chapters 5 and 6 are dedicated to two education systems: Ireland and the French Community of Belgium. Finally, conclusions are drawn following RQ4: is the OECD education policy advice worth following?

A mixed-method design was employed because the analysis needed both qualitative and quantitative evidence to be thorough. This decision reflects accepted practice in evaluation research (Desimone, 2009; Heck, 2004; Stern et al., 2012), but mixed-methods approaches are progressively being adopted as a *de facto* standard procedure in many social sciences (Bryman, 2006).

This chapter uses a narrative style to guide the reader through the research design. This is to keep a consistent flow, to introduce the main concepts or operative definitions only when they are needed, and to provide a clear rationale to each procedure. However, a summary of data collection instruments and analytical methods is provided at the beginning of each section, and a schematic overview of the whole research is available in Appendix 1.

Education policy advice (RQ1, Chapter 3)

Sampling frame	OECD publications, including: PISA international and technical reports, thematic reports, OECD blogs and databases
Sample	All PISA international and technical reports, five issues of *PISA in Focus*, OECD databases and 21 articles from the OECD blogs
Data collection instruments	Systematic online search in OECD websites and databases
Analytical methods	Content analysis (reconstruction), literature review (validation)

Despite the simplicity of the concept (the policy recommendations given by the OECD to interested countries), identifying, summarising and validating the EPA was a challenging process which required addressing multiple issues.

Reconstructing the education policy advice

In the preliminary stages, which are summarised in the theoretical section earlier on in this chapter, a multi-agent practical reasoning model was devised (Figure 2.2). The model was meant to serve as a blueprint: it was hypothesised that most instances of the EPA would take a similar form.

In the main analysis, the model was applied to OECD publications through an iterative two-step process—following Krippendorff (2004)—of (1) unitising and sampling, and (2) categorising / abstracting.

Unitising and sampling

In content analysis, sampling units are 'units [of text] that are distinguished for selective inclusion in an analysis' (Krippendorff, 2004, p. 98). To reconstruct PISA-based education policy advice, the most theoretically-relevant sampling units are the PISA international reports from 2000 to 2012, the monthly *PISA in Focus* series as well as thematic and country reports. Outside its dedicated publications, PISA is referenced in many other OECD products: for example, *Education at a Glance*, the *Education Policy Outlook* series and CERI reports; or online resources like the *OECD Observer*, the *Education & Skills Today* blog and the *Education GPS* database. Overall, the OECD publishes around 250 new titles per year and many of them use PISA data as a source of evidence. To keep the sample size manageable, priority was given to primary sources (the international reports) and to publications that were likely to contain some EPA (such as the *Education Policy Outlook* series). The order of sampling was:

1) International reports
2) *PISA in Focus*
3) *Education Policy Outlook*
4) *Education & Skills Today*
5) In-depth reports (e.g.: Top of the Class, OECD, 2009e)
6) *OECD Observer*
7) *Education GPS*
8) CERI reports

Once the sampling units were decided, two strategies were used to locate the EPA (the coding unit, Krippendorff, 2004).

The first was to focus on the sections most likely to contain policy advice (the context units, Krippendorff, 2004). In international and other reports, this included: the foreword; the executive summary; conclusions to chapters; and specific sections targeting policymakers such as the "Policy Implications" chapter or the "Key policy issues" section, which tend to be a recurrent feature in OECD publications. Online material, instead, was filtered using the "PISA" tag and by searching for contributions by Andreas Schleicher. Eventually, 21 articles were selected, 20 from the period 2011–2013 and one from 2003; most came from *Education & Skills Today* but some came from *OECD Insights*, the *OECD observer* and the *Times Education Supplement* magazine.

The second strategy was to search within whole sampling units for keywords, and in particular for words or phrases indicating deontic modality. Deontic modality 'concerns what is possible, necessary, permissible, or obligatory, given a body of law or a set of moral principles or the like' (von Fintel, 2006, p. 2). Keywords included:

- Modal verbs, and especially *should*, *must*, *might*, *may*.
- Semi-modal verbs: *need to*, *ought to*.
- Verbs used in a deontic sense such as *suggest*, *recommend*, *advise*; and related nouns: *suggestion*, *recommendation*, *advice* (usually by just using the stem *suggest-*, *recommend-* and so on).

This allowed the capture of problem-setting statements like the following:

> 'Education systems with large proportions of students performing below, or even at, Level 1 should be concerned that significant numbers of their students may not be acquiring the necessary literacy knowledge and skills to benefit sufficiently from their educational opportunities.' (OECD & UNESCO-UIS, 2003a, p. 73)

And it also allowed the capturing of actual EPA:

> ‘An important policy objective should therefore be to strengthen the role that education systems play in moderating gender differences in occupational expectations.’ (OECD & UNESCO-UIS, 2003a, p. 146)

It could be argued that this double approach of focusing and scanning might still miss recommendations that are outside the context units or that do not include deontic expressions. This is a fair objection, but it is in the interest of the OECD to produce EPA that is clear, visible and coherent across publications. *Repetita iuvant*, it helps to repeat, and if a hypothetical recommendation appeared only once, in a secondary publication, hidden within the text, then policymakers would hardly notice it and this would imply that it is not considered to be too important by the OECD. Moreover, online resources such as blog posts were filtered by topic but read in their entirety.

Instead of extracting all policy recommendations from all sampling units above and then reorganising them, policy recommendations were categorised into policy areas before moving from one unit in the list to the next. This was done to reduce redundancies, as it had been noticed that OECD publications tends to restate the same or very similar concepts across one another. Consider, for example, the following sentence, which appeared verbatim in five different international reports: ‘as expenditure per student on educational institutions increases, so also does a country’s mean performance’ (OECD & UNESCO-UIS, 2003a, p. 113; OECD, 2001a, p. 93, 2004b, p. 102, 2007b, p. 61, 2010g, p. 34, 2014g, p. 34). By logically organising the EPA before moving onto another unit, instead, it was possible to adopt an incremental approach, whereby new information redefined and fixed previously created categories until saturation was reached.

Categorising and abstracting

There was a distinct feeling that the OECD policy recommendations were meant to be clustered by area, but it was also evident that the OECD had not done so in a systematic fashion. The EPA was never given as one single umbrella strategy, but rather emerged as a collection of recommendations that were spread around several textual units. Therefore, a single OECD programme for education had to be *reconstructed* by iteratively coding and grouping policy recommendations according to their area of intervention. Other researchers had encountered similar difficulties when trying to specify what the OECD suggests that countries should do (Bieber & Martens, 2011).

The first hurdle was understanding what areas of intervention the OECD considers to be related to education (i.e., deciding the categories).

A starting point was provided by the Declaration on Future Educational Policies in the Changing Social and Economic Context (OECD, 1978). Information from the Declaration was complemented by two databases: the *Education Policy Outlook* Reforms Finder (OECD,

2015d), an interactive database containing the policy interventions reviewed in the published reports, and the *Education GPS*, '*the* OECD source for internationally comparable data and analysis on education policies and practices' (OECD, 2016b, emphasis in the text).

Another source used to identify the policy levers potentially addressed in the EPA was the PISA background questionnaires. The selection of the contextual variables underpinning the questionnaires is informed not only by research, but also by policy interests. This happens with all LSAs, from TIMSS and PIRLS (Chrostowski, 2004; Kelly, 2003; Mullis, personal communication on 2 October 2014) to PISA. Therefore, it was thought that OECD recommendations would likely gravitate around background indicators.

Harvey-Beavis (2002) noted that 'in May 1997 the OECD Member countries through the [… PGB] established a set of priorities and their relative individual importance to guide the development of the PISA context questionnaires' (p. 33). The themes captured by background indicators 'had to be of enduring policy relevance and interest' (OECD, 2005b, p. 36); they had to be internationally comparable and 'add significant value beyond that which can be accomplished through national evaluation and analysis' (*ibid*.); they also had to be consistent with other PISA cycles. The 'main goal' of background indicators 'essentially is to guide priority setting and decision making in educational policy', because policymakers 'want to understand why students achieve certain levels of performance' (OECD, 2013c, p. 171).

Background variables included student-, school- and system-level factors such as family SES, gender, migrant status, learning climate, teacher CPD, tracking, government funding or accountability systems in place (see assessment frameworks and technical reports from with Harvey-Beavis, 2002; to Klieme & Kuger, 2014). In general, they tended to map onto the following policy areas of the *Education GPS*: equity, evaluation and quality assurance, finance and funding, learning environment, organisation and governance, school leadership and teachers.

The educational areas covered by the Declaration (OECD, 1978), the *Education GPS* and the background variables became the categories sorting the recommendations extracted from international reports and other OECD publications. The outcome of this exercise was a draft table summarising and organising the education policy advice. The best criterion with which to assess its quality would have been to determine the extent to which the contents of the table actually reflected the views of the OECD. In other words, did the reconstruction adequately capture the education policy advice "construct"?

To test this, drawing from the idea of 'stakeholder-derived theories' (Coryn et al., 2011; Patton, 2008), the draft table was submitted to Andreas Schleicher via email (26 September 2014). The Director of PISA was deemed to be the person in the best position to assess the accuracy and thoroughness of the EPA, given his involvement in its production and

dissemination. Schleicher agreed for the most part that the analytical reconstruction represented the EPA and made some amendments (30 September 2014). The final EPA table can be found in Chapter 3 (Table 3.2).

Validating the education policy advice

Once it was established with the help of Schleicher that the EPA table provided a sufficiently accurate representation of the OECD policy recommendations, the actual advice was validated along three dimensions: internal coherence, consistency and soundness of the programme theory (i.e.: Is it reasonable to expect that following the recommendations would contribute to the realisation of the goals of the OECD?).

The internal coherence is the extent to which different policy directions are not in conflict with one another, and it can be assessed both theoretically and empirically: theoretically, one can envision a situation in which all recommendations were implemented and look for potential breaks in the policy mechanisms; empirically, one can look at existing country policy configurations and see if any of them occur together more often than others.

The theoretical approach drew from research on policy coherence (Nilsson et al., 2012) and findings are presented in Chapter 3. Analytical methods exist also to examine policy configurations (one example is qualitative comparative analysis, see Ragin, 1987), but this research borrowed only their underlying ideas. Specifically, Chapter 5 and 6 highlight the instances in which specific country policies conflict with one another, or when parts of one policy embrace the EPA and other parts contradict it.

The consistency of the EPA was analysed in terms of its stability over time and its connection with PISA findings, under the assumption that the EPA should be data-driven. If policy recommendations were not evidence-based or did not change when PISA data suggested that they should, the OECD branding of PISA as a scientific instrument to inform policymaking could be criticised. The policy advice itself might still be sound and valuable, but then one could argue that if the objective is to acquire policy recommendations, countries could cut costs by not participating to PISA and just following the advice. This analysis was carried out by matching the EPA with co-textual information, charts and data.

Finally, the soundness of the programme theory made explicit by the EPA was evaluated. Chapter 3 will show, however, that the EPA tends to be more of a collection of individual recommendations from different policy areas than one interwoven strategy. Each of these recommendations also subtends a programme theory, and therefore a theory-based evaluation should be carried out for each single recommendation. Experienced evaluators have shown that this is not feasible (Delahais & Toulemonde, 2012; Weiss, 2000).

Consider one actual recommendation, 'Reduce disabled or immigrant students' segregation by better allocating special education teachers where they are most needed'. A validation exercise would firstly have to clarify all constructs implied in the argument, such as "disabled" or "segregation". Then, a programme theory would have to be drafted linking existing teacher allocation practices (there are probably as many as there are countries) to new, 'better' ones, and showing how this would reduce student segregation. Evidence for (and against) all links would have to be produced. The programme theory should be amended in light of new data and eventually the whole 'theory of change' (Weiss, 1997) would have to be evaluated.

Repeating this process for all recommendations is impractical, yet at the same time, the EPA should not be accepted at face value.

A compromise was found by adopting the following strategy. Firstly, evidence was sought that each policy recommendation *could* produce the stated outcomes, but not *how* or *why*. This made the process more manageable while still providing some bases from which to evaluate OECD claims. Secondly, it was recognised that different types of evidence had different weight: studies linking *policies* to educational outcomes weighed more than studies linking policy *outputs* to educational outcomes.

Policy outputs are variations in indicators such as school accountability, degree of tracking in the system, level of parental choice, or teacher self-reported hours of professional development, that are thought to be caused by policies. These indicators are not policy; they are the result of policy and external factors. Almost all secondary analyses of PISA data correlate policy outputs with country achievement but have little to say about policy processes, whereas the assumption underpinning the EPA is that introducing certain policies, whose outputs can be summarised by certain indicators, bring about changes in performance or equity.

Therefore, best evidence was sought in meta-analyses or research syntheses, to increase the probability that the reported effects could apply to a range of widely different interventions in different countries. Articles that were not meta-analyses were reviewed only when they had been cited by the OECD as the foundations of specific recommendations.

The use of education policy advice for national policymaking (RQ2, Chapters 4–6)

Sampling frame	PISA participants
Sample	91 experts from 61 countries (extensive approach) 25 experts from 2 education systems (intensive approach) About 800 documents from 2 education systems (intensive approach)
Data collection instruments	Expert questionnaires (extensive approach)

	Study visits in 2 education systems (intensive approach) Expert interviews (intensive approach) Online and library-based documentary searches (intensive approach)
Analytical methods	Rankings and graphs (extensive approach) Delta-convergence table (intensive approach)

Defining "alignment" to OECD recommendations

Once the EPA was framed for this research, the degrees of connection between the EPA and country policies had to be defined more clearly. Research on PISA has spoken of the 'impact', 'use', 'utilisation', 'effect', 'influence', 'interpretation', 'reception', 'consequences' or 'implications' of PISA to indicate the interactions between PISA and country policies. Depending on the scope of the investigation and the scholarship in which it took place, this interaction may or may not involve the EPA. It could be about media reception, policymakers referencing PISA outcomes to justify reforms, or the emergence OECD values in national policies. Indeed, the literature reviewed in Chapter 1 showed that the most frequent interaction between PISA and country policies is simply the entrance of the names of the OECD and PISA into policy discourses.

However, high media or political attention does not necessarily translate into policy action. This type of interaction—which could be called the "reception" type—has probably the weakest links with country changes in performance or equity (Martens & Niemann, 2013, showed that there is hardly any relationship between a country's ranking and media coverage). Therefore, it is the least useful for evaluation purposes.

A less frequent type of connection is "acknowledgement" of the OECD or PISA in national policies. This happens for example when the introductory statement for a new policy (such as an amended curriculum or a restructuring of the Inspectorate) refers to OECD peer reviews or PISA outcomes as objective evidence that a country's education needs reforming to improve quality or equity, and claims that the policy being introduced will be a solution to the problem. This interaction embraces at the country level the Problematisation or the Goals and Values stages of the practical reasoning model in Figure 2.2. At the same time, it gives credit to (acknowledges the role of) the OECD or PISA for helping the country notice that there was an issue. It is thought to be less frequent than the reception type simply because there are fewer policy documents than there are policy debates.

Sometimes OECD or PISA acknowledgement is accompanied by target-setting. In this case, PISA outcomes are used both to problematise an issue and to measure policy effectiveness at a future stage. However, neither simple acknowledgement nor acknowledgment with target-setting necessarily have strong links with the EPA: even if PISA

results and OECD values are mentioned, there is no guarantee that the policy was actually drafted following OECD recommendations.

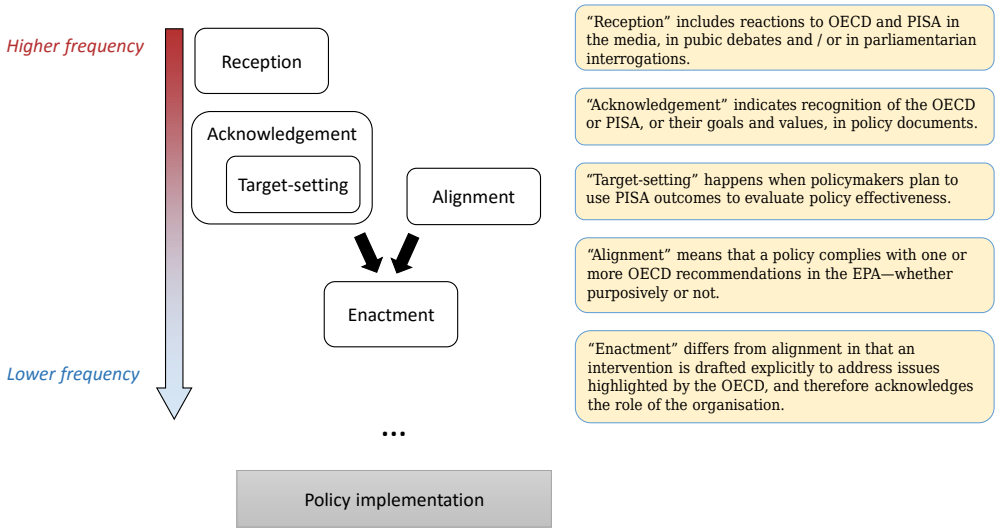
When this happens, the interaction becomes policy "alignment". A policy can be said to be aligned with the EPA when it operationalises the Means–Goal argument in Figure 2.2 whereby 'if countries do *q*, goals will be reached'. An aligned policy purports to do that "*q*". Since there are various recommendations composing the EPA, there are various "*q*s" which a policy may seek to do, and therefore various degrees of policy alignment. Policy alignment is a key component in the evaluation of the EPA: only if positive results are achieved by policies aligned with the EPA can this latter be claimed to be good advice. This was the major shortcoming of previous research on the policy effects of PISA: lacking a theorisation of the EPA, it stopped at the Problematisation or the Goals and Values stages without addressing the Means–Goal argument.

An important point to note is that policies need not acknowledge the OECD or PISA to be aligned with the EPA. For example, the OECD recommends making provisions for early childhood education, but one country's policies on this may be informed by European frameworks or American experiences rather than by the OECD. If a policy that "happens" to be aligned with the EPA were found to be effective, it could still be argued that the EPA is sound even if it was not deliberately followed in that specific instance. Indeed, sound advice should work regardless of whether the advisor is credited for it.

One final type of connection is when a policy is both aligned with the EPA and acknowledges the contribution of the OECD or PISA. These interventions are drafted explicitly to address issues highlighted by the OECD and may be more closely informed by findings from PISA or peer reviews. Given that this is an analogous process, at the country level, to policy enactment at the school level (Ball, Maguire, & Braun, 2012)—whereby a lower-level agent tries to make sense of, and act upon, directions coming from the top—this type of interaction was called "enactment".

The increasing degrees of connection between country policies and OECD recommendations are shown in Figure 2.3.

Figure 2.3: The increasing degrees of connection between country policies and OECD recommendations



Source: Own elaboration

Notice that reception, acknowledgement and alignment pertain to a policy level that could be called "of intentions" and which is separated from policy implementation. In other words, the existence of a policy that purports or demands for a series of actions to be initiated is no guarantee that those who are supposed to initiate them will actually do so. Policy implementation is taken into account in this research because even the strongest case of EPA enactment cannot have any effect if educational practice does not change. A separation of the level of intentions from that of implementation is what Hopkins et al. (2008), drawing from Stake's (1967) evaluation model, defined policy 'intents' and 'observations'.

Measuring policy alignment

Once these operative definitions were set, there was the issue of measuring country interactions with OECD or PISA in over 70 education systems. Following a similar approach to that used by Martens et al. (2010) in their investigation about the influence of international organisations on national policymaking, the analysis was split into two lines of inquiry: one, more extensive, broadly exploring policy alignment with the EPA at the cross-country level (Chapter 4); the other, more intensive, seeking to establish as accurately as possible the extent to which specific national policies in two education systems had been informed by the EPA (Chapters 5 and 6).

Extensive approach (Chapter 4)

For the extensive approach, data on country policies were collected through expert questionnaires. Another method that had been considered was using policy databases from UNESCO (UNESCO-IBE, 2015), the EU (Eurydice, 2016b) and the OECD (OECD, 2015c),

but there were drawbacks. The main one was that these databases provide too few details about a policy to determine its degree of alignment with the EPA.

In a very thorough study, Braga, Checchi, & Meschi (2013) tried to circumvent this issue by considering policy alignment with some external framework at an aggregate level, by the number of policy changes in one same policy area. For instance, if they found that three reforms dealt with the expansion of pre-primary education, each would weigh 1/3 of a unit; the degree of country alignment at one point in time would then be 0, 1/3, 2/3 or 1, depending on the number of interventions that had happened since the baseline year. This is a clever approach, but it has limitations. Firstly, policies tackling the same issue are generally not additive: two policies on teacher autonomy may have been introduced by different governments and have opposite purposes; one may have replaced the other; or the meaning of "autonomy" may have been different each time.

Moreover, not all policies can be easily matched to OECD recommendations. For example, in 1995 the Hungarian parliament voted a resolution 'offering financial incentives to upgrade education and research and improve quality' (Garrouste, 2010, p. 253). While the policy is about resource allocation and quality, the EPA links quality to *equitable* resource allocation. What could an appropriate alignment value be, in this case? 0, 1/2, 1? Or perhaps -1, because we do not know whether the policy was equitable or not.

To avoid incurring in these issues, the approach chosen for this study was to involve experts who could comment on policy alignment with the EPA in their respective countries. The use of experts for data collection was motivated by their effectiveness as a source of information (Martens et al., 2010) and their ability 'to bridge the divide between case studies and the comparison of a large number of countries based on more general and publicly available data' (Dorussen, Lenz, & Blavoukos, 2005, p. 317).

The sampling frame included researchers, academics, governmental and non-governmental officers, and independent consultants. Some were PISA national project managers, members of the expert groups, or they were affiliated with institutions related to the development or administration of PISA; others were academics and researchers in education known by or introduced to this researcher in a snowball fashion.

Overall, a total of 91 experts, representing 61 countries that took part in PISA 2009 and 2012, were contacted via email and asked to give an overall anonymous rating of their country's alignment with OECD recommendations on a 5-point Likert scale ranging from 'Not at all' to 'Completely' (Appendix 3, Table 2.C). Emails were sent in March–April 2014 and a follow-up to all non-respondents was sent in April 2015. In total, 40 answers were received from 30 education systems, which is equivalent to a response rate of 44% representing half of the countries.

Respondents were provided with the opportunity to provide a breakdown of their overall rating by completing an anonymous online questionnaire based on the EPA called the OECD *Policy Advice Implementation Questionnaire* (PAIQ; Aloisi, 2014; a paper version is available in Appendix 3, Table 2.D). The contents of the instrument matched the EPA with the exception of the sections added by Schleicher, which were received only sometime later. The only practical difference is that Recommendation 3.5 in the EPA to 'stimulate a supportive school climate and environment for learning' does not appear in the PAIQ.

Respondents had to tick the OECD recommendations they thought their country had followed, and their answers were sent through an automated email service. The recommendations that were ticked were coded "1"; those which had not been selected were coded "0". Sixteen country breakdowns were collected in this fashion, which resulted in a 16 x 14 matrix of 0s and 1s (Appendix 3, Table 4.N). With this information, it was possible to rank countries in terms of coverage of the EPA and also in terms of the priority each system gave to equity policies.

Intensive approach (Chapters 5 and 6)

For the intensive approach, two education systems were selected as case studies: Ireland and the French Community of Belgium. The inclusion of case studies was motivated by the realist view whereby 'causal outcomes follow from mechanisms acting in contexts' (Pawson & Tilley, 1997, p. 58 emphasis removed). Many sources give credit to this view. Evidence from the literature review highlighted the pivotal role of timing and context when trying to understand the impact of EPA on individual countries. Levin (2005) underlined the importance of origin and adaptation of education reforms. Scheerens et al. (2015) provided evidence of the 'loose coupling' between governmental policies and school practice, whereby high-level policies are not necessarily implemented in the classroom. The importance of context in educational change was recognised even by the OECD (Snyder, 2013), despite its more mainstream narrative about the possibilities of cross-country comparisons.

All education systems are potentially interesting and many, as Chapter 4 shows, have followed the EPA to a certain extent. The choice eventually fell on Ireland and French-speaking Belgium after securing the opportunity to obtain key data on site. Two research visits were organised at the Educational Research Centre of Dublin, in Ireland (www.erc.ie), and the service for the Analysis of Systems and Practices in Education of the University of Liège, Belgium (http://www.aspe.ulg.ac.be). To ensure methodological consistency between the two visits, each followed the same research protocol.

Six case study questions were used to guide the within-country analysis. They reproduced, on a smaller scale and at a higher level of detail, the four general research questions framing this thesis:

CSQ1) What information emerges about country performance from international and national PISA reports, and what policy advice is given?

CSQ2) What were the main educational policies in this country between 1995 and 2009?

CSQ3) Reception and use:

a. What were the reactions to OECD reports, in terms of reception or future planning?

b. What are the contact points between the EPA for this country and actual interventions?

CSQ4) What is the context surrounding these interventions—i.e., what other influences concurred to shape these policies?

CSQ5) What evidence exists about the plausible impact of policies (are they theoretically sound, were they implemented, what are the recorded effects)?

CSQ6) To what extent was following the EPA useful for this country (the benefits outdid the costs, relevant new knowledge was produced, cognitive or non-cognitive outcomes, or equity, improved…)?

The first four case study questions (CSQ1–4) deal with the production and use of the EPA (RQ1–2), whereas CSQ5 and 6 informed the analysis on the plausible influence of EPA-informed policies on country scores (RQ3, see the next section in this chapter).

CSQ1 maps onto RQ1 and it serves to differentiate *general* EPA from *country-specific* EPA. Individual countries are frequently cited within international PISA report as an example of good or bad practice. If the 'naming and shaming' theory is accurate (Carvalho & Costa, 2014; Lingard, 2011; Takayama, 2008), policymakers will be more likely to look at the specific recommendations in the text surrounding the parts in which their country is mentioned. However, most countries also produce national versions of the international reports, and these may be the only source policymakers get to read. Therefore, an enquiry into country alignment with the EPA should firstly check that the EPA provided in the national reports is itself in line with the EPA provided by the international PISA reports.

Answering CSQ2 was necessary to create a policy database, whereas CSQ3 helped to discriminate reception and acknowledgement from actual alignment. CSQ4 maps onto both RQ2 and RQ3: on the one hand, it seeks to trace policy origin and adoption (Levin, 2005) as well as the extent to which the OECD *did not* influence national policymaking (RQ2); on the other hand, it serves to understand whether similar policies in different contexts produced different educational outcomes (RQ3).

Two sources of evidence were used in each case study: policy documents and education experts. Policy documents were consulted to identify the main educational interventions in the

period under consideration (CSQ2), understand their alignment with national or international EPA (CSQ3) and outline their context (CSQ4). Education experts were interviewed to answer all case study questions.

The process of data collection and analysis was iterative, because the data provided by either source could complement, contradict or point to data from the other, which led to further data gathering, comparison and analysis. It also served for cross-validating the data: policy experts validated this author's interpretation of the findings from the documentary analysis, but knowledge of the documents meant that experts' opinions did not have to be taken at face value.

Policy documents consisted of parliamentary acts, guidelines and circulars, as well as official policies developed by governmental institutions such as educational boards, authorities, regulators and inspectorates. To qualify for inclusion, policies had to satisfy two requirements. Firstly, they had to be introduced within the period under consideration (1995–2009); this excluded bills, green papers, and any *proposed* change in the education system, but strategies, plans and other frameworks were still examined as they often led to the identification of specific policies. Secondly, they had to affect, at least on paper, a large share of their target population; this excluded pilot studies but not policies targeted at a small part of the total student population such as, for example, those aimed at the bottom 10% of achievers. The data were gathered from academic, institutional or professional websites (e.g., teacher unions) and libraries.

Policies are both text and discourse (Ball, 1993); therefore, the analysis of policy documents was complemented by interviews with educational experts, who could enrich the written narrative with personal accounts or informed insider judgements. Educational experts were identified during the documentary analysis and by asking researchers in the host institutions. They generally included MoE officers, academics, researchers, members of professional associations and teachers' unions officials.

Once the experts had been identified, they were contacted via email or telephone. If they expressed an interest in being involved in the study, a face-to-face interview was arranged following a strict interviewing protocol that sought to minimise ethical issues (see the "Ethics Considerations" section, below) and maximise effectiveness and consistency between the two case studies; its format drew from relevant literature on interviewing methods (Rowley, 2012), as well as from research on carrying out expert interviews (Aberbach & Rockman, 2002) for evaluation purposes (Gugiu & Rodríguez-Campos, 2007), and it is outlined below. Overall, 16 experts were interviewed following this protocol in Ireland, and 9 in Belgium.

An initial letter 'giving the name and status of the researcher carrying out the study, a brief rationale of the study including its purpose and value and why the individual [was] being

invited to take part' (ESRC, 2015, p. 42) was sent. Upon acceptance, a topic for the interview (generally one or two policies) was agreed. This was done to increase the quality of the data, as sometimes over 15 years had passed since the policy under consideration had been introduced. Focusing on few policies was a way to give participants the opportunity to prepare for the interview by reviewing old documents and recalling their involvement at the time. Nevertheless, during the actual interviews some experts still preferred to give an overview of the issues surrounding many policies rather than a detailed account of one or two.

Interviews were semi-structured, in the sense that they comprised a fixed part using CSQ1–6 but additional questions could be asked to follow-up important points (Kvale, 2007). Following the recommendations of experienced evaluators and established practice (Bryman, 2012; Patton, 2002), all 25 interviews were meant to be recorded, but eventually only 9 of them could be because, on the one hand, some experts agreed to only 'have an informal chat' about their topic, whereas others asked to turn off the recording equipment before beginning the interview.

It is acknowledged that some details may have been lost when the interviews were not recorded, but using note-taking and no audio recording forced both parties to focus only on key information and moved the emphasis from the interviewee to the policies. The interviewee was signalled when direct quotations were being written down. At the end of each interview, all points covered were recapitulated and direct quotations re-read, which provided the opportunity for further discussion, clarification and collaborative construction of meaning.

After the interview, the notes were integrated with information from the documentary analysis and a draft was sent to the interviewees, with their contribution highlighted, for further review. This step was needed to ensure anonymity, but it carried the risk that participants could disagree with the research outcomes (Boddy et al., 2010). This issue was addressed by allowing informants to comment only on the parts of the document to which they had contributed. The reviewed drafts flowed into a core document reporting, for each individual policy: a description, the country context and national influences, the contact points with education policy advice, policy implementation and plausible impact. Chapters 5 and 6 are based on these core documents.

Finally, country alignment with the EPA was summarised drawing from the concept of "delta-convergence", that is, 'the decreasing distance of policies towards an exemplary model, for example a model promoted by an international organisation' (Holzinger et al., 2008, p. 83; see also Jakobi & Teltemann, 2009). The delta-convergence value is a useful summary measure of the degree of EPA-alignment, and it has already found applications in research on PISA (Bieber & Martens, 2011). However, it was difficult to assign justifiable numerical values to Belgian and Irish convergence, because the 'exemplary model' provided by the EPA

is not sufficiently precise to be interpreted in a quantitative fashion. Therefore, the principle of delta-convergence was adapted to produce a visual representation of the progressive alignment of national policies to the EPA.

The plausible influence of EPA-informed national policies on country achievement and equity (RQ3)

Sampling frame	PISA participants (extensive approach) Education policies 1995–2006 from 2 education systems (intensive approach)
Sample	57 countries with at least 3 data points, but variable (extensive approach) About 50 policies from 2 education systems (intensive approach)
Data collection instruments	OECD, World Bank, UNESCO databases (extensive approach) Study visits in 2 education systems (intensive approach) Expert interviews (intensive approach) Online and library-based documentary searches (intensive approach)
Analytical methods	Correlation analysis (both Pearson's and Spearman's), linear regression and multilevel growth modelling (extensive approach) Willms' (2006) framework, graphs and review of existing evaluations (intensive approach)

The extensive and intensive strategies used for measuring policy alignment were also used in the last part of the analysis, because RQ3 follows directly from RQ2. The issue when looking for policy effect was how to link some *specific* national policies, via policy outputs, to educational outcomes, and draw valid conclusions. Extant literature provided direction but no individual publication could solve the problem. This section outlines the nature of the problem and then introduces the approach used in this research.

Some strategies for capturing policy effects and their challenges

Many studies using data from international LSAs, and especially econometric analyses, sought to find a relationship between some country-, school- or student-level indicators and student achievement (Bol, Witschge, Van de Werfhorst, & Dronkers, 2014; Fuchs & Wößmann, 2007; Hanushek & Wößmann, 2011; Wößmann, Lüdemann, Schütz, & West, 2007). However, these studies are generally not designed to account for the contribution of specific policies in producing the values the indicators take.

What this literature refers to as "policies" are in fact policy outputs (Schlicht-Schmälzle, Teltemann, & Windzio, 2011; Teltemann, 2014)—an observed variation in some indicators over time that is assumed to be due to policies. Policy outputs are often interpreted as static systemic features rather than the dynamic result of governmental interventions. Therefore, even studies linking output indicators (such as grade retention) to specific interventions (e.g., economic incentives for teaching SEN students, Goos et al., 2013), tend to rely on cross-

sectional data. Actual policy change on outputs was analysed by Jakobi & Teltemann (2009), but the link between outputs and country performance was outside the scope of their investigation.

A key issue when linking national policies to policy outputs to educational outcomes is accounting for time. Each moment is associated with a certain socio-political and economic context which may or may not enable the implementation of EPA-informed policies (Steiner-Khamsi, 2010). 'Different temporal conditions occur for different kinds of institutional experiences', and therefore, 'the correct amount of social goods and services [should] be delivered at the appropriate point in the social process for the integration, maintenance, and improvement of the system to which policies are directed' (Hayden, 2006, p. 10).

The first problem when dealing with time is deciding how long it takes from the moment a policy is implemented to the moment it affects students enough for changes in learning trajectories to be detectable. The word "deciding" was used on purpose, as it is apparent from the literature that this is mostly a matter of a researcher's judgement.

Fullan (2000) argued that school districts can improve in 8 years—or even in half that time (Fullan, 2007, p. 18)—but he did not set any time scale for systemic changes. "Success stories" such as Ontario's were presented as evidence that sizeable change can be achieved rapidly (Levin, Glaze, & Fullan, 2008). Likewise, Earl, Watson, & Torrance (2002) argued that the English National Strategies for Literacy and Numeracy 'made significant changes in primary education throughout England in a remarkably short period of time' (two years), but also that they were 'not yet deep-seated enough to continue without consistent pressure and support' (p. 35). In fact, follow-ups showed that the improvements may have been the result of score inflation due to test familiarity (Tymms, 2004), and that they plateaued anyway after four years (Hargreaves & Fink, 2006).

Borman et al. (2003) found that sustained implementation of comprehensive school reform models increased achievement, and that after five years the effect sizes were twice as large as at the baseline. Harris (2011) argued that implementing successful reforms can take ten years—an estimate agreed by Jakobi & Teltemann (2009)—whereas Wedell (2009) stated: 'it is clear that the successful implementation of educational change takes a long time'—perhaps a generation or more (p. 18).

The generational timescale was adopted by Goos et al. (2013). While gathering lagged data on systemic features that may have affected PISA results, the authors noted that 'deviations of merely a few years [between observations of a same variable] were neglected' because 'the implementation of reforms […] most often occurs over the course of several years or decades' (p. 65). This viewpoint was considered unsatisfactory for this research, as it

assumes that policy effects are almost time-invariant. If that were the assumption, why did the author decide to collect data from different periods at all?

For the purpose of this research, it was decided to focus on policies introduced between 1995 and 2009. This time frame was informed by the literature above and was a compromise between scale and data availability. The further back in time one goes, the harder it is to find policy documents and receive accurate information from educational experts. Yet this time frame allows policy effects to become manifest after up to about 10–12 years, though in this case only pre-2000 reforms could be associate to PISA outcomes (recall that PISA was administered in 2000, 2003, 2006, 2009 and 2012).

The second problem when dealing with time is deciding how to model it. Common approaches in evaluation include: instrumental variables, whose variation causes variation in the predictor but not in the outcome; regression discontinuity, whereby the results of students just above and below an policy-induced cut-off point are compared; or differences-in-differences (or double difference), where data between two groups are compared before and after a policy has targeted either one (Khandker, Koolwal, & Samad, 2009; Schlotter, Schwerdt, & Wößmann, 2011).

Theoretically, these strategies should be able to trace student outcomes back to policy, but the myriad confounding factors and assumptions involved means that, in practice, they might not be more accurate than expert judgement (Meghir & Rivkin, 2010). Consider for instance an extensively-studied policy that is not part of the EPA: changes in class size. The effectiveness of changing class sizes has been explored using a wide range of methods, including those outlined above (Hanushek & Wößmann, 2011; Meghir & Rivkin, 2010; Schlotter et al., 2011); and yet it took seventy years for research to reach some consensus (in one country only, the United States) that reducing class size has a small positive impact, and the debate is still ongoing on the matter of cost-benefits (Chingos, 2013).

Time series models, such as autoregressive and moving average models (Shumway & Stoffer, 2011), were also considered because they can accurately model trends. However, when one considers that a country participating in all PISA cycles has at most five data points for its outcome variable (country estimates of student achievement) and that missing data are common in these kind of large-scale datasets (Jakobi & Teltemann, 2009), an important gap appears once again between what is theoretically recommended and what is practically viable.

The few data points available for the outcome variable are related to a wider issue: whether PISA is an appropriate measure to detect policy effects at all. This issue is tackled in Chapter 4 but was recognised from the early stages of the design of this research. The problem might not be PISA *per se*, but the moderate effects of educational interventions. Despite the powerful

narrative about the possibilities for systemic change (Barber & Mourshed, 2007; Fullan, 2009; Hargreaves, 2009; Schleicher, 2013a; Sugrue, 2008), in fact:

> 'The empirical evidence on the impact of reform in these five settings [governance, curriculum, assessment and accountability, inspection, privatisation and choice] is quite limited, and almost non-existent in some cases […], especially taking into account the enormous effort involved. They have changed some relationships in important ways, including giving individual schools much more autonomy in some settings. They have changed the work of administrators in significant ways, and certainly intensified pressures on teachers. They have increased the importance of parents in a number of ways. They appear on the whole to have had small impacts on student achievement levels' (Levin, 2005, p. 189).

Similarly, Scheerens et al. (2015) found that PISA, rather than confirming the importance of education policies for student outcomes, suggests that 'contextual, "ecological" conditions […and] school composition in particular' (p. 32) strongly limit the capability of policy to make a sizeable difference, at least in the short period: 'Compared to the effect of SES, the impact of variables that are prone to manipulation by policy makers is quite modest […;] our findings may be a warning against over-optimistic expectations of educational policy measures' (p. 33).

Given all these methodological limitations, the approach followed in this study is described below.

Extensive approach (Chapter 4)

The extensive approach was designed in three steps that used increasingly targeted data to capture policy effects.

The first step was to analyse trends in performance and equity since PISA started in 2000. If the EPA is adopted, implemented and effective, then some increases in performance or equity should be observed. Score and inequality changes over time were modelled using correlational analysis, individual country regressions and multilevel growth models.

Multilevel or hierarchical modelling (Gelman & Hill, 2007; Goldstein, 2011; Raudenbush & Bryk, 2002) is a type of regression analysis that can simultaneously model country-specific and average trends by taking into account between- and within-country similarities. To do this, it assigns a variance component not only to Level-1 units (within-country scores over time), but also to average country achievement (the Level-2 units). Evidence of policy effects was sought in country trajectories, in the clustering of OECD and partner countries, and in trends in the within- and between-country variance. Detailed information on the modelling approach, including variables and diagnostics, are available in Appendix 5 and 6.

The second step was to contrast the differential effects of three policy outputs (economic productivity, equitability and access to pre-primary education) on PISA outcomes. All three

outputs are known to correlate to PISA outcomes, but it was posited that the factor that is more malleable by specific education policies (access to pre-primary education), would show a stronger relationship with changes in PISA outcomes. Smaller correlations were expected for the "generic" educational factor (equitability) and even smaller for the non-educational one (productivity). By comparing the size of the three indicators, the second step attempted to set an empirical limit on the extent to which education reforms could be expected to produce measurable changes in country outcomes.

The third step used information collected through the expert questionnaires (see "Measuring policy alignment", above) to relate country achievement and equity to their degree of alignment with the EPA. Because of the small sample size for which there were data, it was decided not to use multilevel modelling at this stage. Instead, countries were ranked twice: by degree of alignment with the EPA, and by their ability to translate policy statements into action. This ability was measured through a "government effectiveness" indicator, reflecting 'perceptions of the quality of public services, the quality of the civil service and the degree of its independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies' (Kaufmann, Kraay, & Mastruzzi, 2010, p. 7).

The weighting of policy alignment ranks by government effectiveness ranks produced an overall index of policy alignment with OECD recommendations, or "EPA Index". The EPA Index was correlated (using Spearman's ρ) to country changes in performance and equity to analyse the relationship between policy alignment and outcomes.

To summarise, the first step modelled educational outcomes in performance and equity; the second related outputs to outcomes; and the third linked policy alignment with outcomes. Altogether, they sought to gather increasingly targeted evidence that country alignment with OECD recommendations may affect their PISA outcomes.

Intensive approach (Chapters 5 and 6)

Further evidence was gathered during the two study visits (guided by CSQ5 and CSQ6). Irish and Belgian policies were analysed using Willms' (2006) LSA-based policy evaluation framework. The framework divides education policies into categories: 'universal' interventions seek to raise the achievement of all students; 'SES-targeted' policies intervene on the performance of students with a low socio-economic status (SES); whereas 'performance-targeted' interventions try to improve the results of low achievers. According to the OECD, policy effectiveness can be inferred by analysing variations on indicators that are specific to each category.

In this research, such variations were interpreted in light of findings from existing policy evaluations carried out by national experts. Existing evaluations were preferred to original analyses for the same reason underpinning the use of expert interviews: it was thought that the collective knowledge produced by local researchers and research teams would exceed by far in breadth, depth and rigour the possibilities of someone from outside the system and with very limited time. This also allowed for critiquing the ability of Willms' framework to detect policy effects on PISA.

Ethics considerations

Given the involvement of external informants in the research, ethical considerations were addressed following general guidelines (BERA, 2011) as well as departmental ones. A critical point was to ensure confidentiality and anonymity (Boddy et al., 2010): because of the relatively small pool from which experts could be drawn and the use of snowball sampling, participants were likely to know each other. While, on the one hand, this made it easier to recruit new informants, it also meant that each interviewee was highly identifiable by just mentioning his/her affiliation. Therefore, several measures were put in place to protect the participants' right to anonymity:

1) Preliminary ethical approval was sought from the School of Education's Research Ethics and Data Protection Sub-Committee.
2) Upon contact, a letter outlining the research project and the reason for expert participation was sent (Appendix 4).
3) Prior to the interview, participants were asked to read and sign an informed consent form (Appendix 4).
4) Once the interviewee's comments had been integrated into the draft analysis, the part of the document informed by the interview was sent to them for further review. This was to ensure that the expert felt he/she could not be identified by other parties.
5) Care was taken in the final report and in this thesis to remove critical information and to identify all interviewees by a code ("I" or "B", for Ireland and Belgium, followed by a number). In some occasions, it was felt that even this system may compromise the anonymity of the participant, in which case the number was not shown.

Chapter 3: The OECD education policy advice (EPA)

This chapter is articulated in the following parts:

- A reconstruction of the OECD education policy advice.
- An analysis of its coherence and internal consistency, including the extent to which it is informed by PISA.
- An analysis of its soundness focused on recommendations aimed to improve country equity, quality of provision and student performance, as well as their use of educational time.

It is argued that the EPA is highly consistent over time, partly because it is somewhat disconnected from PISA data, but it is not conceptualised as a coherent programme. Some recommendations find strong backing in the literature, whereas others do not, suggesting that the fragmented identity of the OECD has an impact on its ability to promote a uniform educational strategy.

Introduction

The difficulty in reconstructing the EPA is that it has been forty years since the OECD last took an explicit official stance on what educational objectives countries should pursue. The OECD "Declaration on Future Educational Policies in the Changing Social and Economic Context" (OECD, 1978) focused on:

- student acquisition of basic competences 'as conceived within each country' (PISA did not exist);
- the development of schools as a place that contributed to 'self-reliance, sense of responsibility and co-operative spirit of young people';
- teacher training and professional development to establish links 'between the school and adult life'; gender equity; provisions for migrant students and other disadvantaged groups;
- ensuring that 'any necessary procedures related to educational choice, assessment and certification' did not compromise pupil development;
- finding the correct balance between general and vocational qualifications, and the provision of work experience during schooling; and lifelong learning.

Since then, only two formal recommendations on education were made, both in 2005: one, on earthquake safety in schools (C(2005)24), and the other on cross-border higher education (C(2005)147). This contrasts dramatically with the 40 Recommendations and 4 Declarations

on environment; the 9 Decisions, 12 Recommendations and 1 Declaration on chemicals; or the 1997 Convention on Combating Bribery in International Business Transactions (OECD, 2015b). At least until PISA, there were few collective attempts to regulate education compared to other sectors.

There are grounds to claim that PISA provided an opportunity to spell out educational objectives and policy advice more explicitly. One objective of the INES programme—and PISA by extension—was to 'promote sound education policy development in member and partner countries by highlighting successful educational outcomes and strategies' (OECD, 2012d, p. 4). As argued in Chapter 1, if PISA were just a 'thermometer', it would only pair high costs with minimal returns to feedback, like a school test used for a summative purpose (Hattie & Timperley, 2007; citing Lorrie A. Shepard et al., 1995).

Instead, it is because PISA data are analysed in a broad policy context (OECD, 1999c) and linked to certain recommendations that PISA has a prominence which IEA assessments currently do not enjoy. The EPA is a strong and increasingly relevant feature of PISA reports: 'initial reports issued by the OECD […] have been expanded from a single, mostly descriptive book reporting on PISA 2000 to the five-volume report on PISA 2009 published in 2010' (Klieme & Kuger, 2014, p. 8).

The OECD education policy advice

Chapter 2 described how a blueprint to capture policy recommendations in OECD publications was developed and applied. Firstly, a working definition of "education policy advice" was set and a multi-agent practical reasoning model was devised to discriminate policy statements from other types of text. Then, relevant OECD publications were sampled and, in an iterative fashion, the EPA was identified, categorised and abstracted into a general framework. The categories were based on PISA background indicators, as well as on the areas and themes in the *Education Policy Outlook* Reforms Finder (OECD, 2015d; also Appendix 3, Table 3.A) and in *Education GPS* (OECD, 2016b; also Appendix 3, Table 3.B).

These sources provided an effective starting point for organising many policy recommendations, but not all. For instance, the following excerpt was initially coded as "equity – inclusion – SES":

> 'Some inclusive policies try to reduce between-school socio-economic segregation by redrawing school catchment boundaries, amalgamating schools, or by creating magnet schools in low-income areas.' (OECD, 2010e, p. 105)

After accounting for other sources, however, it was noticed that school amalgamation was an example of policies on student stratification and grouping (a new category); redrawing catchment boundaries was an example of school choice (another new category); whereas the

formation of magnet schools was not an actual policy option, as the OECD claimed that in fact more effective solutions existed (OECD, 2013e, p. 59).

In many other instances, the policy categories used by the OECD were inadequate to frame the EPA. The excerpts in Table 3.1 all come from a single page in the 2012 international report (OECD, 2014g, p. 253):

Table 3.1: Instances of EPA that were difficult to categorise using only OECD policy categories

EPA	Issue
Some countries 'have implemented targeted policies to improve the performance of low-performing schools or students'.	Policies targeting low achievement may be concerned with both equity and accountability or represent a category on their own.
'Some countries […] have given schools and local authorities more autonomy but have recognised that autonomy works only in the context of collaboration and accountability'.	Should autonomy and accountability be a single category?
Some countries 'have reshaped the organisation of schools to facilitate collaboration and economies of scale between individual schools by creating school clusters'.	School collaboration and clustering are somewhat captured by the "organisation" theme, but they are not measured by any indicator.
'Recognising that a positive learning environment is key to promoting positive attitudes among students which, in turn, promote learning, [… some countries] have improved their students' attitudes, dispositions and self-beliefs towards school in general, and towards mathematics in particular, by, for example, reforming their curricula so that they are better aligned with students' interests and 21st century skills'.	Here, learning environments and student attitudes are invoked to introduce (with a rather sudden logical leap) curricular reforms. Curriculum is a sensitive topic for the OECD, because much of the novelty of PISA is built around its purpose to measure "skills for life" rather than curricular achievement. Therefore, since the beginning no indicator was dedicated to this aspect (see for instance Harvey-Beavis, 2002; OECD, 1999a). Yet, the issue of curricular alignment with 'students' interests and 21st century skills' frequently appears in international reports and other publications.

To summarise, the way the OECD categorises education policies had to be further developed to include all policy recommendations in the sampled publications. As mentioned in Chapter 2, a new framework was produced and submitted to Andreas Schleicher to review. The final framework, which integrates his comments, is shown in Table 3.2:

Table 3.2: the OECD education policy advice

AREA / RECOMMENDATION	SAMPLE POLICIES
1. Educational Time	
1.1. Increase overall time spent in education	Policies expanding access to pre-primary education, targeting enrolment or changing the age of entry into school.
1.2. Increase instruction time	Interventions targeting lateness, attendance, by introducing full-day schooling, or by increasing the time spent in classes.
2. Equity	
2.1. Reduce disabled or immigrant students' segregation	Interventions reviewing the sorting of pupils in institutions, or by better allocating special education teachers where they are most needed.
2.2. Reduce the gender gap	For instance, by promoting subjects/career paths that are traditionally dominated by one gender to boys or girls.
2.3. Reduce the streaming / grouping / stratification of students	By reviewing the academic or vocational paths, reducing the transfer of students to different schools, reviewing retention policies or ability grouping within classes. *To prevent drop out, ensure high quality vocational education and training and equivalence and transferability between vocational education and training and academic pathways*.
2.4. Reduce school competition	By changing school incentives to take in the 'best' students or by tackling the elite status of private institutions. *In school choice environments, mitigate its negative impact on segregation with options such as controlled choice schemes, incentives to make disadvantaged students attractive to high quality schools, school selection mechanisms and vouchers or tax credits. Improve disadvantaged families' access to information about schools and to support them in making informed choices.*
2.5. Help socio-economically deprived students	By providing additional instruction, free learning resources, fee waivers, cash transfers/vouchers or a modified curriculum.
3. Quality of provision	
3.1. Strengthen the teaching profession	Policies focusing on: qualifications required, initial teacher training, allocation of excellent teachers/principals where they are needed the most, retention of teachers in the most challenging schools, teacher pay, in-service training/professional development, career paths.
3.2. Allocate resources more efficiently or in a more equitable way	By changing the share of spending towards public and private schools, by targeting socio-economic depressed areas or acting upon the evaluation of intervention programmes. *Available resources and the way they are spent influence students' learning opportunities. Use funding strategies, such as weighted funding formula, that take into consideration that the instructional costs of disadvantaged students may be higher*.
3.3. Pair more autonomy with accountability measures	For instance, by increasing school autonomy in budget, curriculum and/or assessments **while at the same time** holding schools accountable for their results by public posting of results, external monitoring of standards (both aspects are required for the association to have a positive effect).
3.4. Promote cooperation between parents, teachers, principals and schools	Through collaborative programmes, by increasing parental involvement in school matters and in school governance, by facilitating/incentivising school-based research or the sharing of good practice.
3.5. Stimulate a supportive school climate and environment for learning	*Low performing disadvantaged schools are at risk of difficult environments for learning. Policies specific for these schools need focus more than other schools on the following: prioritise the development of positive teacher-student and peer relationships; promote the use of data information systems for school diagnosis to identify struggling students and factors of learning disruptions;*

	adequate student counselling, mentoring to support students and smoother their transitions to continue in education. In addition, these schools may benefit from alternative organisation of learning time, including the duration of the school week or year, and in terms of the size of schools. In schools [there] can be a policy to reinforce student-student and student-teacher interactions and better learning strategies.
4. Student Performance	
4.1. Help low-performing students	Through feedback and appraisals for students, or by establishing early-warning mechanisms and providing a modified curriculum or additional instructional support for struggling students.
4.2. Standardise the educational outcomes	For example, by introducing central, external exit examinations for qualification purposes.
4.3. Review curricular priorities	By introducing curricula that are 'better aligned with students' interests and 21st century skills' (OECD, 2014g, p. 253), or by examining curricular choices 'in the light of the performance of other countries' (OECD 2007a, p. 113)

Sources: Own elaboration drawing from OECD (2001a, 2004a, 2012c, 2013d, 2013e, 2013f, 2013g, 2014g, 2015c, 2004b, 2007a, 2010b, 2010c, 2010d, 2010e, 2010f, 2010g); OECD & UNESCO-UIS (2003a, 2003b); Schleicher (2011b, 2012, 2013a, 2013b, 2013c, 2013d, 2013e, 2013f). The parts in italic were added by Andreas Schleicher.

In this new organisation, the EPA is divided into 4 areas: educational time, equity, quality of provision and student performance. The first category was included to capture recommendations on pre-primary education and instruction time. The fourth category differs from the third in that it includes policies targeting student performance more directly.

Clearly, any attempt to frame the EPA includes an element of judgement, and other researchers may have come to different conceptualisations. For instance, the recommendation 4.2 ‘Standardise the educational outcomes’ could be merged with the recommendation 3.3 on accountability, as centralised external exit examinations (CEEEs) are often use for this purpose. While the econometric literature considers both CEEEs and public posting of school results as accountability measures, the OECD tends to associate accountability only to public posting of results, whereas its position about CEEEs is more ambiguous. Therefore, the two policy levers were kept separate in the framework.

There are some redundancies and small inconsistencies among the sample policies. Surely this is due in part to the fact that Schleicher could only have a brief look to the document, but they also testify the difficulty of organising policy recommendations that have yet to be officially formalised in a coherent system. This is why the table is presented with its flaws, with the acknowledgement that it could and should be improved by researchers and especially by the OECD.

Validating the OECD education policy advice

The purpose of the EPA is to provide countries with ways of developing policies pursuing the OECD objectives of performance, or 'productivity' (OECD, 2013c, p. 171), efficiency and equity (regardless of the extent to which individual members subscribe to these values). Within the framework of practical argumentation, validating the advice is an attempt to answer this question: Is it reasonable to expect that following the recommendations would contribute to the realisation of the goals of the OECD?

Two strategies were employed to assess this: firstly, that the EPA does not contradict itself across categories (internal coherence) or over time (consistency); secondly, that the EPA is supported by at least some empirical evidence.

Internal coherence

On a broad level, it is clear that OECD recommendations are organised along some core dimensions that tend to be more complementary than conflicting. There is a core interest towards inclusion, lifelong learning, teacher training and structural institutional arrangements. The issue is when one moves away from general topics and analyses the organisation of more specific policy options. Policy areas crop up and disappear across the publications informing the framework in Table 3.2, so that the end result looks more like a collection of stand-alone suggestions than a coherent strategy.

For instance, the OECD is undoubtedly interested in the governance of schools and education systems, yet it is harder to understand how "governance" is defined, what policy levers would affect educational governance, and how policies for governance are supposed to fit with other policies.

According to the *Education Policy Outlook* Reforms Finder, governance policies affect 'education priorities' and the 'organisation of decision-making process'. In the *Education GPS*, in contrast, governance is divided into 'parental involvement', 'school choice', 'public & private stakeholders' and 'school autonomy'. In the PISA 2006 international report, these four policy levers appear in the same chapter, though the chapter itself not called "governance" but 'school and system characteristics', and it includes also non-governance policies such as 'admittance, selection and grouping policies' and 'school resources' (OECD, 2007b, p. 7). The 2006 report does mention 'school founding and governance', but they fall under 'public and private school management'. The technical report from the same administration (OECD, 2009c) never mentions the word "governance" in the main text.

Similar critiques can be made for other policy areas. The outcome is a policy vision for education that has some but not complete cohesion, it is a 'patchwork of loosely connected

topics [… that] have been strung together to provide a minimum semblance of coherence' (Henry et al., 2001, p. 52). The OECD often notes that some policies are more suited to some contexts than others and that recommendations are just 'options' (OECD, 2013d, p. 16); it also notes that policies may be 'applied in combination' (*ibid.*) but does not commit to explaining how.

In summary, there *is* one such thing as the OECD education policy advice; however, much of it remains latent and there are great disparities in how frequently and how clearly recommendations belonging to different categories appear in OECD products.

To some extent, this is unavoidable because of the complexity of education as a subject matter. On top of this, the fragmentation of the OECD as an organisation and the relative novelty of education as an independent and valuable branch mean that there might not be a unique view on what policy areas countries are supposed to address and how. The OECD has been called a 'network of networks' (Carroll & Kellow, 2011, p. 64) and it appears to be cyclically affected by some sort of identity crisis: 'Over the past 20 years or so, the diagnosis has been the same: The OECD is in search of a distinct identity […]. No wonder, therefore, that the two earlier secretaries-general have had problems in explaining what [… it] actually does' (Marcussen & Trondal, 2011).

If one considers that the recommendations in Table 3.2 are a synthesis of hundreds of pages and they do not reach policymakers in such a structured layout, it should not surprise that there are varying national level interpretations of the "common" goals and values of the OECD.

This thesis is not alone in claiming that more should be done in terms of organisational clarity of the EPA. After the first PISA administration, the Strategic Development Group asked for three improvements (OECD, 2003c, pp. 10–11):

1) 'A better conceptionalisation [*sic*] and operationalisation of the key policy objectives that are driving the development of PISA';
2) 'The establishment of key policy levers that shape these policy objectives';
3) 'The development of analytic approaches on the ways in which the policy levers interact with the policy objectives'.

Hopkins, Pennock, & Ritzen (2008) noted that even though PISA should 'help countries and economies to develop insights into what kinds of good instructional practices, school organisation and system level decisions make a difference' (p. 75), 'the policy drivers that may or may not be associated with the performance of individual countries have not yet been subject to *systematic* scrutiny' (*ibid.*, emphasis added). The PGB recognised this issue in 2005,

but suggested that the earliest that policy levers could be feasibly investigated would be with PISA 2012 (OECD, 2005a). As of 2011, this was still work in progress (OECD, 2011e).

Consistency over time

The consistency of policies over time is related to their credibility (Majone, 1996): if the EPA kept changing, the OECD could not aspire to be a stable point of reference for country policies. At the same time, the EPA is supposed to be informed by PISA findings, and therefore it should change only when empirical evidence points to other directions. How does the OECD deal with this tension? In general, the OECD seem to privilege consistency over accuracy, and what follows is evidence in support of this impression.

Firstly, there is a certain disconnect between PISA evidence and policy recommendations. The OECD tends to reuse previously-given recommendations even when the indicators change. While 70 indicators were created between 2000 and 2009, more than 55 were used just once (OECD, 2013c). Of these 55, some were domain-specific indicators which were meant to be excluded when the domain changed. Other variables, however, were simply dropped. If policy recommendations were highly dependent on these variables they might have changed as the variables did, but this is not the case. Many recommendations are repeated verbatim in every international report, regardless of the presence or absence of evidence from PISA to support them.

For instance, it was affirmed in 2003 and 2006 that 'for some countries' improving quality and equity 'may mean taking measures to […] improve early childhood education' (OECD, 2004b, p. 198, 2007b, p. 205). This is a rather general statement ('improve' in what sense?), and the only empirical support given to it by PISA was that student attendance at pre-primary education correlated to PISA outcomes in 2003 (no data were available in 2006).

Notice that being able to report a relationship between student responses in a self-completion questionnaire on pre-primary attendance and PISA outcomes 10 years later is still an exciting finding (it was 'striking', OECD, 2004b, p. 267). However, the OECD did not include evidence about the specific value of being able to participate in early childhood education in the general recommendation quoted above, in either in 2003 or in 2006. It was as if PISA evidence was not necessary to drive the policy point, even though the recommendation was made in PISA reports.

Secondly, the EPA is often informed by case studies rather than PISA data, even in the PISA reports. This is exemplified by the growing tendency of including "success stories" in dedicated boxes within the main analysis, and using them to give policy recommendations. There were no such boxes between 2000 and 2006, 6 in 2009 and 18 in 2012. These case

studies often have only a loose connection with PISA or the work of the OECD, even though they are meant to be exemplars of what is possible to achieve by following the EPA.

Consider the case of Brazil. Brazil 'offers a good example of how low-performing countries can use international benchmarking to improve their education systems' (OECD, 2010d, p. 102). In both 2009 and 2012, the OECD mentioned that Brazil improved in PISA and proceeded to enumerate a series of national interventions in education (see OECD, 2014g, pp. 76–78). Unsurprisingly, all these reforms were praised. This was regardless of whether they had been conceived before or after PISA, or whether they had been implemented or were still at the planning stage. For instance, the OECD reported that an examination system for teacher certification had been delayed but at the same time—in contradiction to this fact—that its establishment sent 'a strong signal of what content and pedagogical orientation should be developed' (OECD, 2014g, p. 78).

Sometimes the praise did not even take into account whether Brazilian reforms actually represented the views expressed by the OECD in other circumstances. A positive comment was made about an 'important increase in education expenditure' (p. 77), although the OECD has often stressed precisely the fact that higher spending does not necessarily translate into better results: 'the Slovak Republic, which spends around USD 53 000 per student, performs at the same level as the United States, which spends over USD 115 000 per student' (p. 34). Interestingly, since 2005 Brazil has spent about the same amount per student as the Slovak Republic, yet the former was praised for its increase in commitment whereas the latter for its parsimony.

Overall, the boxes dedicated to Brazil were informed by a range of sources, but the links with PISA were restricted to a few outcome figures, even if this was supposedly 'a good example' of the usefulness of PISA as an instrument for policymaking (OECD, 2010d, p. 102).

The consistency over time of the EPA and the use of external sources to inform it raise questions about the extent to which "PISA-based" policy recommendations are actually based on PISA. This is problematic, as it goes to the heart of the knowledge-for-policy construction process of the OECD. If the OECD does not always distinguish what was shown empirically from what was assumed, it is making the same instrumental use of PISA as countries.

Does the education policy advice precede or follow PISA evidence?

Even when policy recommendations are linked to PISA evidence, it can be shown that PISA data is sometimes presented and interpreted to suit pre-existing policy stances. Three such practices are identified below.

First practice: unidirectional reading of graphs

The first practice is interpreting what are essentially flat graphs in the most favourable sense. Consider a regression line of equity on income distribution leaving 96% of the variance unexplained (OECD, 2010e, p. 32). The OECD suggests that this weak relationship 'shows that equity in educational opportunities can be achieved even where income is distributed highly inequitably' (*ibid*.).

This is correct, but so are the following equivalent readings: where income is distributed highly inequitably, equity in educational opportunities is often *not* achieved; countries can have high inequality in income distribution regardless of their equity; or, below-average income inequality can be achieved independently of the educational equity of the system. These are all paraphrases of the same concept (no links between equity and income distribution were observed), but the OECD chose the phrasing that could better represent 'what is possible' in education (McGaw, 2008, p. 223; OECD, 2010e, pp. 3, 18, 2013h, p. 4; Schleicher, 2009a, p. 100), rather than what is happening.

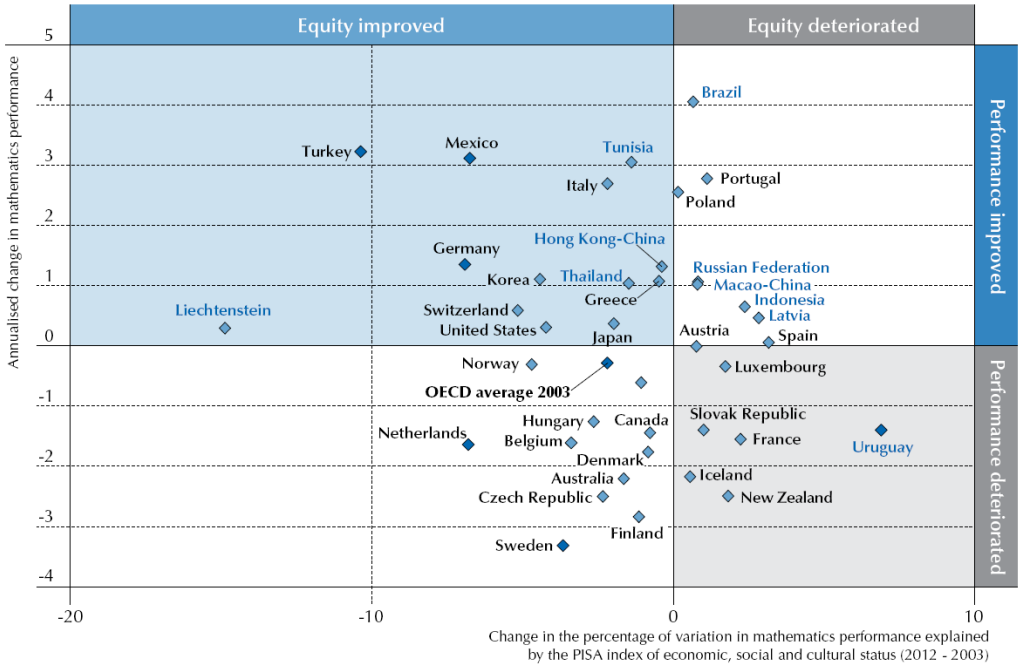
Another example of this "narrative of possibility": 'the findings from PISA suggest that systems prioritising higher teachers' salaries over smaller classes tend to perform better, which corresponds with research showing that raising teacher quality is a more effective route to improved student outcomes than creating smaller classes' (OECD, 2010f, p. 106).

The counterargument is as follows. Leaving aside the problem that a causal link between teacher salary and teacher quality has yet to be established (Springer & Balch, 2009; Tirivayi, Maasen van den Brink, & Groot, 2013), using information from the same table (OECD, 2010f, p. 85, Figure IV.3.7) one could equally claim that countries pairing small classes and/or low teachers' salaries with high expenditures tend to perform better and, in many cases, to be more equitable. Or that 4 out of the top 5 performers have either larger classes, higher salaries and low expenditures (Hong Kong, Shanghai); or smaller classes, lower salaries and high expenditures (Canada, Finland).

Second practice: confusing graphs

A second practice is to display both statistically significant and non-significant observations in the same picture, with only a slight change in shading to separate the two. This can increase the perception of strong bivariate relationships when there is little evidence for them. An example is a scatterplot of changes in equity and performance in the second volume of the 2012 report (OECD, 2013e, p. 57, reproduced below), in which the point representing the statistically non-significant change in Liechtenstein may add to the impression of a negative correlation. The OECD uses this graph to argue that 'improvements in performance need not come at the expense of equity' (OECD, 2013e, p. 57).

Figure 3.1: Change between 2003 and 2012 in the strength of the socio-economic gradient and annualised mathematics performance



If one separates significant and non-significant changes (Appendix 3, Figure 3.A and Figure 3.B), another narrative that could be told is that, firstly, in the vast majority of cases, variations in equity or performance (or both) between 2003 and 2012 were so small that the chances are they were not different from zero. Secondly, even if they were different, they would be uncorrelated, as the regression line in Figure 3.A shows. Therefore, one could claim that improvements in equity were not beneficial for country performance.

In fairness, in the few instances in which changes in equity and performance reached statistical significance, a negative correlation can indeed be observed: those countries that managed to reduce the association between student socio-economic status and achievement by 10 percentage points in 10 years also experienced a 25-point increase in scores. However, this correlation is computed on observations from only six countries and one aggregate value: the OECD average—which, by the way, indicates that in OECD countries an increase in equity between 2003 and 2012 was associated with a *decrease* in PISA scores.

Murphy (2014) expressed similar concerns and concluded:

> 'inconsistent use of statistical methods in its 2006 analysis and selective reporting of its 2009 analysis, and the fact that over both surveys attention has been drawn to positive findings […] while negative results for consequential accountability policies have been hidden or played down leave PISA and the OECD open to criticism that their work does not meet the highest standards of objectivity and favours certain schooling policies.' (D. Murphy, 2014, p. 913)

Third practice: variable *p*-values

A third related practice is that the OECD is sometimes flexible in terms of what confidence level counts as being statistically significant. *p*-values are some of the most widely misinterpreted statistical concepts (Gelman, 2013; Schervish, 1996), but the OECD provides an excellent definition: 'differences are labelled as statistically significant when a difference of that size, smaller or larger, would be observed less than 5% of the time, if there were actually no difference in corresponding population values' (OECD, 2010f, p. 139).

Unfortunately, the accuracy in the definition is paired with some laxity in the practice. 10% levels are regularly reported (though clearly labelled), and sometimes indications of statistical significance are left to the end-of-chapter notes. For instance, the main text may report that 'countries that use standards-based external examinations tend to perform higher, even when accounting for national income' (OECD, 2010f, p. 46). One has to read the endnotes to find out that the confidence level for that finding was 7% because of the 'large amount of missing data regarding standards-based external examinations' (pp. 59–60).

Together, the three practices highlighted here reflect a general attitude whereby PISA data is used to confirm rather than challenge pre-existing assumptions. This is somewhat unavoidable, given the correlational nature of PISA and the political role of the OECD, but privileging consistency over accuracy means unlinking the EPA from the evidence: it creates a vicious circle in which, at some point, the EPA precedes (and survives) the data. At that point, it is unclear what other value PISA may have rather than being used as a mere accountability stick and a rich source of data for researchers.

Recent trends

It should be acknowledged that in recent years the OECD has substantively increased its grounding of background variables in research, which might progressively affect the content of its recommendations.

For instance, PISA returns information on the students' engagement and sense of belonging to school. The rationale for this is that positive attitudes can influence achievement and lifelong learning. In 2000, the OECD wrote: 'The evidence is […] that those who do like school perform better than those who do not' (OECD, 2001a, p. 108). It immediately proceeded to qualify this statement by recognising that the observed relationship could be the outcome of reversed causality, but still provided the following "PISA-based" recommendation: 'the results suggest that school policy and practice should [… create] an engaging learning environment for all students' (*ibid.*).

In 2003, the OECD deferred an investigation on the relationship between attitudes towards school and performance, but repeated both the recommendation and the argument's warrant that affectionate students 'are more likely to do well in their studies' (OECD & UNESCO-

UIS, 2003a, p. 122). The investigation came out in the same year. The foreword mentioned once again that students' participation and sense of belonging are important 'because of their relationship with student learning' and 'because they represent a disposition towards schooling and life-long learning' (OECD, 2003a, p. 3). In the analysis, however, the author noted that there was no observable correlation between engagement and performance and wrote: 'It cannot be infered [*sic*] from these findings that efforts to increase student engagement are likely to lead to better literacy skills' (Willms, 2003, p. 29). In other words, for two administrations the OECD had given "evidence-based" policy suggestions which were not supported by evidence, but by a policy stance that presumed a link between engagement and performance.

The 2012 international report saw a major change: the mechanisms explaining *how* engagement was believed to affect student learning were based in research and clearly articulated. As a result, the sense of belonging was no longer theoretically conceptualised as directly related to performance. Therefore, when it was found that 'the relationship between students' sense of belonging and their mathematics performance was weak in 2003 in all countries and economies, and remained weak in 2012' (OECD, 2013f, p. 55), this did not shake the general argument: engagement fosters other aspects such as truancy, teacher-student relations and school climate, which in turn impact on the whole school community. Put in these terms, even if the EPA was not particularly different from that in 2000, it was more coherent and provided better bases with which to evaluate its soundness.

Soundness

This last example may indicate a new direction in the way policy issues are integrated into the design of PISA, one that sees research evidence play a much more prominent role. This is summarised by the following excerpt from the 2012 assessment framework:

> 'Given that PISA is entering its fifth cycle of surveys, it is time to reconsider the overarching framework that guides the development of questionnaires and the way that this information is used in analysing and reporting data. […] Questionnaire content has largely been determined by contemporary priorities rather than long-term policy and research goals. […] PISA has matured. As a sound and stable basis for international comparative studies and trend information on education systems, PISA requires a coherent architecture for future context questionnaires that balances stability with innovation, and domain-specific with general issues.' (OECD, 2013c, p. 168)

In the 2012 international report, clear relationships between variables were made manifest and, as a result, so were the mechanisms leading to performance or equity outcomes. The draft framework for 2015 (Klieme & Kuger, 2014) continues along the same lines, meaning that the EPA might become more coherent in the future. This would be a welcome change from an

evaluation perspective, because the OECD policy recommendations thus far have been characterised by an almost complete lack of an explicit programme theory. As was claimed in Chapter 1, the OECD writes what countries should do, but not how or at least why its advice should work, with the exception of having been attempted by some successful countries.

This is a major shortcoming of the EPA which affects the credibility of the OECD's knowledge-for-policy construction model. To state it more overtly: if evidence from PISA—the instrument conceived to shape or at least support policy recommendations—has had a marginal role in shaping or supporting policy recommendations; and if a coherent explanatory framework for measuring and interpreting measures was only recently devised; what has the OECD been basing its policy advice on? And how does the OECD know that following what it suggests will produce any positive effects?

In the remainder of this chapter, evidence in support of the EPA is sought, with a proviso. Each recommendation (e.g., changing retention policies or establishing additional support for struggling students) involves a complex but unstated network of causal mechanisms linking interventions with outcomes. It would be impossible develop a programme theory to evaluate each of them, and some evaluative approaches (e.g. contribution analysis, see Mayne, 2012) require even more than a "simple" programme theory. What was possible, rather, was to select from the four main educational areas (educational time, equity, quality of provision and performance, see Table 3.2) a few recommendations that were emblematic of a specific issue in terms of the way the OECD grounds and delivers advice, and to review their overall plausibility, drawing from education literature.

The recommendations analysed below deal with pre-primary education (educational time); gender, streaming and school choice (equity); teacher professional development and school accountability (quality of provision) and curricular changes (performance). Problematic aspects are summarised and general conclusions are drawn at the end of the chapter.

Educational time

The policy recommendation analysed in this section is about extending pre-primary education by expanding access or extending its duration.

Pre-primary education

The OECD advises that 'one of the most efficient educational strategies for governments is to invest early[10] and all the way up to upper secondary' (OECD, 2012c, p. 9).

[10] Although "early childhood" is 'considered to start at conception and continue through the transition to primary school, which may occur as early as age six or as late as age eight' (Engle, Gamer, & Vargas-Barón, 2011, p. 8), in OECD publications "early" education refers to pre-primary education, which officially [Source: UNESCO] starts at ages 3–5 until 6–7.

This policy recommendation stems from interdisciplinary studies (such as Ramey, Ramey, & Lanzi, 2006; Ramey & Ramey, 1998) on the impact of early year interventions on life-span development. This work has recently been reframed by James Heckman as skill formation, whereby early interventions have a higher rate of return than later ones because 'skill begets skill and motivation begets motivation' (key recent publications include Cunha & Heckman, 2010; Heckman, 2008; Kautz, Heckman, Diris, ter Weel, & Borghans, 2014; see also Salverda, 2011, for a summary of the model). Literature from the United States finds ample evidence that some pre-primary programmes are a cost-effective way to address inequalities and achievement (Almond & Currie, 2010; W. S. Barnett, 2011; Camilli, Vargas, Sharon, & Barnett, 2010; Deming, 2009; Duncan, Ludwig, & Magnuson, 2010; Schweinhart & Fulcher-Dawson, 2009; Schweinhart et al., 2005; Wong, Cook, Barnett, & Jung, 2008).

At the individual level, 'striking disparities in what children know and can do are evident well before they enter [primary school]' (Committee on Integrating the Science of Early Childhood Development, 2000, p. 5). At the country level, the policy issue is whether increasing access in pre-primary education could by itself raise country performance (as was the case, historically, with the expansion of primary and secondary education), or whether the focus should be on increasing its quality. Brazil was recently put under pressure to choose between expanding and improving the system after a World Bank evaluation showed that 'children who attend low-quality pre-schools perform the same on literacy tests two years later as do children who attend no pre-school at all' (Evans & Kosec, 2012, p. xi). The authors concluded that 'the quality of a pre-school (and not merely attendance) determines student outcomes' (p.5).

It has been argued that increased access through publicly-funded pre-school programmes may produce 'substantial educational, social, and economic benefits, *but only* if the investments are in programs in which teaching is highly effective' (Pianta, Barnett, Burchinal, & Thornburg, 2009, p. 51, emphasis added). In developing countries, successful early interventions appear to require the 'integration of health, nutrition, education, social, and economic development, and collaboration between governmental agencies and civil society' (Engle et al., 2007, p. 234).

At the same time, it was shown earlier in this document that, although the OECD speaks about quality and coverage in its EPA, in fact it measures only the latter. Drawing from PISA 2009, the OECD (2011b) reported that 'students who had attended *some* pre-primary school outperformed students who had not' (p. 1, emphasis added), and this was 'even after accounting for their socio-economic backgrounds' (*ibid*.). The analysis was expanded with PISA 2012, when most countries were found to have either low rates of enrolment in pre-primary education, or a higher relative risk in terms of performance associated with non-

attendance (or both—see OECD, 2013e, p. 101). This suggests that expanding access to pre-primary education may have an impact on country scores by either slightly affecting the performance of a large number of students, or substantially affecting that of a minority.

The rationale for greater participation finds both theoretical backing (Darragh, 2007) and empirical support. Mitchell (2011) describes the benefits of a governmental strategy to increase participation in pre-primary education in New Zealand. Similarly, the introduction of universal pre-school education in Oklahoma was reported to have improved the school readiness of children (Gormley, Gayer, Phillips, & Dawson, 2005). After evaluating the impact of a pre-school initiative in New Mexico over four years, Hustedt et al. (2009) write that, even though aspects concerning the quality of the programme could be improved, 'continued expansion of the […] initiative is warranted' (p. 6).

Moving beyond the 'quantity versus quality' debate, the availability, duration and quality of pre-primary education have all consistently been found to be related to achievement and equity, even by longitudinal studies (Sylva, 2014). And yet, surprisingly, this has never been considered a key policy issue by the PGB. In 1997, it was completely absent from the list of policy priorities guiding the development of the context questionnaires (Harvey-Beavis, 2002), and no indicators for this variable were included in the 2000 or 2006 administrations. The factor was given low priority by the PGB even in the 2015 cycle and therefore, despite the extensive evidence presented by the Expert Group in favour of its salience, detailed questions about the pre-primary experiences of PISA students were relegated to an optional parental questionnaire (Klieme & Kuger, 2014).

Equity

Half of the OECD educational efforts are dedicated to convincing member countries that equality of opportunities and unbiased outcomes are worthy targets that pay off in terms of economic and social benefits. Therefore, one would expect the EPA covering these aspects to be supported by the strongest evidence. This section analyses recommendations on gender gap, on stratification policies and on school choice.

Gender gap

Perhaps the most unexpected finding about gender gap concerns the extent to which results from international LSAs reflect actual underlying differences. These results generally show a large advantage of girls in reading, a smaller advantage of boys in mathematics and very little difference in science (Eurydice, 2009b). Some researchers (Lafontaine & Monseur, 2009; Lietz, 2006a, 2006b), however, have noted that the differences observed since the early 1990s may in part be explained by methodological changes in LSAs.

This and other evidence led some authors to go as far as to conclude that 'boys are doing as well as girls in reading achievement and have done so for over the past century' (Loerke, 2012, p. 267). Similarly, Corbett, Hill, & St. Rose (2008) reported that the gender gap in the US was marginal and had been narrowing over the previous 35 years. It has often been pointed out that socio-economic inequalities have a much greater impact than gender differences on outcomes, and should therefore be granted higher policy priority (Connelly, Sullivan, & Jerrim, 2014; Corbett et al., 2008; Eurydice, 2009b; Loerke, 2012).

This is not a sufficient reason to neglect the gender issue, though. Female disadvantage in mathematics, albeit small, has been consistently observed both in TIMSS and in PISA (Else-Quest, Hyde, & Linn, 2010), but also in other cross-sectional studies (Dickerson, McIntosh, & Valente, 2015) as well as in large-scale longitudinal studies (Robinson & Lubienski, 2011). Recent research has pointed to a greater variability of male achievement at the extreme ends of the distribution: the performances of high-achieving boys are substantively higher than the girls' in mathematics and science, and substantively lower than the girls' in reading (Baye & Monseur, 2016).

Within countries, boys' and girls' performances also appear to be related: 'In countries where girls lag less behind boys in mathematics and science, they also are more ahead of boys in reading', and vice versa (van Langen, Bosker, & Dekkers, 2006, p. 172). Marks (2008) claimed that this could be capturing the effectiveness of education policies targeting girls, but other researchers have shown that the correlation becomes negative for high performers: at the higher end of the spectrum, the disadvantage of boys in reading decreases and that of girls in mathematics increases (Stoet & Geary, 2013). This is not what one would expect from policies targeting girls, unless they were not effective with the strongest achievers.

The interest of the OECD in addressing the gender gap is certainly justified when it comes to higher education. In the majority of OECD countries, participation and attainment in higher education favours females over males, and yet gender segregation by subject is still evident and stable (Vincent-Lancrin, 2008). This affects STEM subjects (science, technology, engineering and mathematics) but also other traditionally male-dominated programmes like doctoral medicine (S. L. Morgan, Gelbgiser, & Weeden, 2013).

The determinants of the gender gap in secondary and higher education are not fully understood. Once differences in cognitive abilities are discounted (Spelke, 2005; Spinath, Eckert, & Steinmayr, 2014), the remaining theories can be sorted into two interrelated categories (Buchmann, DiPrete, & McDaniel, 2007; Eurydice, 2009b): individual factors (socio-emotional skills, behaviour, attitudes, self-concept) and environmental factors (social expectations including teacher and parents, school bias including resources, curriculum and assessment). Research has uncovered some mutually reinforcing dynamics.

On the one hand, boys and girls seem to have slightly different social and behavioural skills in early years which may contribute to widening the gender gap over time (DiPrete & Jennings, 2012). Specifically, self-concept, self-efficacy, attitudes and motivation towards gendered subjects develop early, affect achievement and alter study and career choices (Jiang, 2014; Perez-Felkner, McDonald, Schneider, & Grogan, 2012). In a meta-analysis of the relationship between intelligence, personality, motivation and gender differences in performance, Spinath, Eckert, & Steinmayr (2014) concluded that 'girls are somewhat better adapted to today's school environment than boys, and this can partially explain why they often outperform boys in academic contexts' (p. 239).

On the other hand, external factors such as peers and teachers can moderate or reinforce inequalities. Peer culture, for instance, shapes masculine identity and may therefore foster or help to prevent antisocial behaviours, which in turn are linked to underachievement and dropout (Ferreira, Santos, Fonseca, & Haase, 2007; Legewie & DiPrete, 2012). Gender-bias in teacher ratings (Campbell, 2015; Hartley & Sutton, 2013; Hinnant, O'Brien, & Ghazarian, 2009; Raag et al., 2011; Robinson & Lubienski, 2011; Robinson-Cimpian, Lubienski, Ganley, & Copur-Gencturk, 2014) influences teacher expectations, student beliefs and achievement.

Changing educational pathways today provide women with 'growing opportunities for pursuing non-STEM degrees and careers that are equally prestigious and socially important' (Mann & DiPrete, 2013, p. 7), and this could explain why gender segregation has remained stable in STEM subjects. At the same time, however, extrinsic motivation has been found as the greatest distinguishing factor between girls that choose to enter a degree in mathematics or physics and those that do not (Mujtaba & Reiss, 2013; Reiss, 2013).

Overall, given the multitude of factors contributing to the gender gap both in secondary school and in the transition to higher education, the OECD recommendation of de-gendering career paths is appropriate, but perhaps insufficient. Surely, following the EPA could mean improving some segregating apprenticeship programmes (Beck, Fuller, & Unwin, 2006) or changing the advice people in an influential position, such as teachers, may give (S. Brown, Ortiz-Nuñez, & Taylor, 2011; Reiss, 2013). However, it could do little to alter the shortcomings of career guidance initiatives (Bimrose, 2009) or to tackle the structural systemic inequalities giving rise to the unrealistic expectations held by many students (Jerrim, 2014).

Following recent trends in its approach to the construction of the EPA, in its latest publications the OECD has begun to address gender equality in greater depth. The new policy recommendations were still somewhat simplistic, but they were clear, targeted and consistent with research findings: give students a greater choice in what they read, privilege homework over video gaming, train teachers to be aware of their own gender biases, build girls' self-confidence, and provide better guidance (OECD, 2015i).

Stratification policies: tracking, grouping, retention

Despite some terminological uncertainties (Betts, 2011), and despite a few outlying findings that are likely due to methodological approaches (Schlicht, Stadelmann-Steffen, & Freitag, 2010; Waldinger, 2006) the sorting of students into separate groups or curricular tracks is generally considered to be ineffective. Over the years, evidence has accumulated from both cross-country studies (Ammermueller, 2013; Haahr, Nielsen, Hansen, & Jakobsen, 2005; Hanushek & Wößmann, 2011; Horn, 2009; Link, 2012; Van de Werfhorst & Mijs, 2010), and national follow-ups to LSAs like those in Germany (Maaz, Trautwein, Lüdtke, & Baumert, 2008), in the Netherlands and Flanders (Prokic-Breuer & Dronkers, 2012), in Hungary (Horn, 2013), and in other countries (see the examples in Salverda, 2011).

Duru-Bellat & Suchaut (2005) concisely explained that 'any restriction on enrolment or early setting aside of certain students, any grouping by distinct level or streaming […] and all phenomena of segregation among schools […] tend to increase social inequality in performance without improving average level or even elite level' (p. 192). Even gifted students might not necessarily benefit from differential instruction, which has been reported in some cases to have a negative impact on their self-concept and anxiety (M. Becker et al., 2014).

By improving labour-market allocation while increasing inequalities (Bol & van de Werfhorst, 2013), tracking reinforces the role of education systems as person-sorting machines (Spring, 1976), which is the opposite of the emancipatory role valued by the OECD. Tracking 'by and large, reinforces [the effects of] family background. It increases inequality and reduces intergenerational mobility for most measures of educational attainment and early labour market outcomes' (Brunello & Checchi, 2007, p. 847). When it comes to grade retention, the academic consensus is even more definite: 'it would be difficult to find another educational practice on which the evidence is so unequivocally negative' (Hattie, 2008, p. 98; citing Holmes, 1989; see also Education Endowment Foundation, 2016).

Among other suggestions, the OECD recommends avoiding the practice of transferring students between schools as a disciplinary action to manage behavioural or performance issues. But the way the EPA is presented resembles more a passing comment than a real point for action.

Firstly, the OECD writes: 'In school systems where it is more common to transfer weak or disruptive students out of a school, performance and equity both tend to be lower. Individual schools that make more use of transfers also perform worse in some countries' (OECD, 2010b, p. 15). It seems that the OECD considers these policies ineffective, an impression corroborate by the following: 'Schools and teachers in [… successful] systems do not allow struggling

students to fail; they do not make them repeat a grade, they do not transfer them to other schools' (OECD, 2012f, p. 4).

But then the OECD adds: 'Transferring pupils who do badly may be partly a symptom, rather than a cause, of schools and school systems that are not producing satisfactory results, especially for lower-achieving students. It is worth noting that the schools with lower transfer rates tend to have greater autonomy and other means of addressing these challenges' (OECD, 2010b, p. 15). Suddenly the issue of student transfer feels like a pretext to introduce another recommendation: increase school autonomy.

Besides, while there is extensive literature on the causes and consequences of school mobility (Gruman, Harachi, Abbott, Catalano, & Fleming, 2008; Mehana & Reynolds, 2004; Rumberger, 2003), and there are some policy evaluations of interventions to reduce 'voluntary' transfers (i.e, those initiated on the students' side, for example Fleming, Harachi, Catalano, Haggerty, & Abbott, 2001), much less information is available on student transfers as a disciplinary policy. Some research suggests that some schools let go (Riehl, 1999) or actively transfer students to boost their performance, but this has negative consequences on the achievement and dropout rates of transferred students (Bowditch, 1993; Rumberger & Palardy, 2005).

On the matter of dropout, Schleicher proposed (somewhat in line with other OECD policy recommendations in this area, OECD, 2011d) a higher degree of compatibility between general and vocational training. This might make the vocational pathway more appealing (or less unappealing), but consider that the most important dropout risk factors originate at the family level and are socio-economic in nature (Freeman & Simonsen, 2015; Rumberger & Lim, 2008). In some countries, students progressing in vocational tracks have very practical issues such as lower access to funding (Farías & Sevilla, 2015).

Increasing student opportunities to move between tracks does not mean that families would actually be able to seize them, or even that they could see such opportunities as available to them (A. Daly & Thomas, 2008). There is a well-known process of self-limitation whereby students 'exclude many institutions that are located beyond perceived boundaries of physical, academic or social space' (Hoelscher, Hayward, Ertl, & Dunbar- Goddet, 2008, p. 145; and see Chapter 6 on the French Community of Belgium). Students who have already dropped out are even harder to reach (Polesel, 2010). Finally, policies trying to bridge vocational and general pathways may be very ineffective. An example is the English Diploma: introduced in 2008, it suffered from a 'low uptake, low understanding, low recognition and high complexity syndrome' (Hodgson & Spours, 2010). It was terminated in 2013.

School competition and choice: school vouchers

The theme of student mobility is closely related to parental choice in school selection and school competition to attract students. This is a very large and controversial research field because it is built upon ever-going debates on public versus private education, or social democracy versus liberal democracy. In this section, only one of the policy options available to countries is examined: school vouchers, which are predominantly used in the US but are increasingly becoming popular in other OECD countries as well.

To give a sense of the partisanship dominating the issue, two definitions follow. According to the Friedman Foundation for Educational Choice, which is in favour of vouchers, 'Vouchers give parents all or a portion of the public funding set aside for their children's education to choose private schools that best fit their learning needs' (Friedman Foundation for Educational Choice, 2015); according to Raise Your Hand Texas, which is against vouchers, 'Vouchers are taxpayer-funded subsidies for private schools with no accountability and restrictive admissions policies' (Raise Your Hand Texas, 2015).

In fact, the landscape is much more complex. Vouchers are 'inherently a flexible instrument that can be designed in many ways' (Hoxby, 2003, p. 14) and because of this, research attempting to find benefits (or flaws) that are common to all interventions has consistently reported mixed results.

In the early 2000s, Hoxby (2003) argued that evidence from the 'first generation' of such interventions was positive. Voucher programmes were successfully targeting disadvantaged or underachieving students who then improved their performance, while also managing to increase the "productivity" (i.e., the 'achievement per dollar spent', p. 17) of public schools (because low-performing students had moved to private schools thanks to the vouchers). Hoxby's (2003) argument and findings were immediately rebutted by Ladd (2003) on the basis of a different theoretical stance and interpretation of the evidence. In the same years, Belfield & Levin (2002, 2005) argued that the debate could not be settled as long as the evidence was being sought in support of pre-existing ideological positions, but they agreed with Ladd (2003) that the size of the observed effects was generally small.

The final blow to Hoxby's (2003) argument was provided by a review by Rouse & Barrow (2009), who wrote: 'The best research to date finds relatively small achievement gains for students offered education vouchers, most of which are not statistically different from zero. What little evidence exists about the likely impact of a large-scale voucher program on the students who remain in the public schools is inconsistent at best' (Rouse & Barrow, 2009, p. 37).

Subsequent research revealed that many claims about the benefits of voucher programmes were based on findings from a very specific intervention (the Milwaukee Parental Choice

Program), which had a small impact and on reading scores only (What Works Clearinghouse, 2012). Generally, research from both the US and other countries reported little or no effect on achievement, and mostly negative effects on segregation (DeLuca & Dayton, 2009; Usher & Kober, 2011; Waslander, Pater, & van der Weide, 2010; Witte, 2009). Their cost-effectiveness was similarly questioned (Yeh, 2010). There is some evidence, however, that these programmes might increase graduation rates and parental satisfaction (Usher & Kober, 2011).

It has been claimed that vouchers increase the productiveness of those public schools experiencing drops in student enrolment because students with vouchers prefer neighbouring private schools (Egalite, 2013a), but methodological variations across studies mean that this last point is far from settled (Egalite, 2013b; Merrifield, 2013). Figlio & Hart (2014) used the introduction in 2001 of the Florida A+ policy to argue that vouchers can have a positive competitive effect, whereas Bowen & Trivitt (2014) used the discontinuation of that same policy in 2006 to argue the opposite.

Finally, it has been found that voucher students leaving the programme to return to public schools initially experience increases in achievement, although the reasons why this might be the case are not understood (Carlson, Cowen, & Fleming, 2013).

Overall, investing in voucher programmes does not appear to be a particularly solid recommendation. In his review of the EPA presented earlier in this chapter, Schleicher mentioned vouchers as an instrument to make disadvantaged students more attractive for high quality schools. This evokes the idea of a "progressive voucher scheme", which can be found in a few specific OECD publications on the topic. On top of funding based on low achievement, progressive vouchers give extra funding to disadvantaged students, which should act both as a means to provide more resources to those who need them most, and as an incentive for schools to take in more low-SES students (Musset, 2012; OECD, 2012c). However, the difference between progressive and "standard" vouchers is not made explicit in PISA international reports, and it is not evident that implementing progressive vouchers would decrease segregation.

Quality of provision

The OECD is very attentive to the institutional features potentially increasing teaching and learning quality. Two recommendations are tackled in this section: measures to strengthen the teaching profession and accountability systems.

Teacher professional development

Teachers can make a substantial difference to student learning (Hattie, 2008; T. J. Kane, McCaffrey, & Staiger, 2013; Muijs et al., 2014), and the OECD is active in collecting data on teachers not only through PISA but also, and especially, through the five-yearly Teaching And

Learning International Survey (TALIS). Among the many policies for strengthening the teaching profession, one specific intervention is discussed here: continuing professional development (CPD), otherwise called "in-service training". The most comprehensive resources on the OECD's stance on CPD are TALIS publications (e.g., OECD, 2009a), thematic reviews such as *Teachers Matter* (OECD, 2005c, 2011g) and a report from an international summit (Schleicher, 2011a).

The OECD policy advice is that CPD should be based not only on teachers' needs, but also on their performance and on the needs of their school. To align these three components, the OECD suggests that teacher profiles could be developed. Teacher profiles lay out professional expectations in terms of content knowledge, skills and effectiveness; they should be outcome-based, informed by research and developed with the help of the category, and should act as 'a framework to guide initial teacher education, teacher certification, teachers' ongoing professional development and career advancement' (OECD, 2011g, p. 11).

Some CPD time should be fixed, some should be based on incentives and some should depend on school needs, and the CPD provided could be partly free and partly a paying service. CPD should be part of the school evaluation system and should be expected from older teachers, too. 'Effective professional development needs to be on-going, include training, practice and feedback, and provide adequate time and follow-up support. Successful programs involve teachers in learning activities that are similar to those they will use with their students' (Schleicher, 2011a, p. 19). Finally, the OECD recommends promoting frequent and structured teacher cooperation to enhance student achievement.

The best evidence available on CPD mostly agrees with the OECD recommendations, especially in terms of CPD having to be structured and systematic. It also adds some further insights. Helen Timperley (Timperley, Wilson, Barrar, & Fung, 2007; Timperley, 2008) confirms that the conditions for teacher learning are similar to those for student learning: existing practice and assumptions should be challenged, new practice should be modelled and teachers should be given the opportunity and the support to apply new knowledge and skills. Opportunities for CPD could be modelled around the indications of current research, for instance the educational effectiveness strand, which has been found to have a greater impact on teacher skills and student outcomes than the holistic/reflective approach (Antoniou & Kyriakides, 2011).

Teacher learning opportunities and skill development should focus on student outcomes; in other words, 'success needs to be defined not in terms of teacher mastery of new strategies but in terms of the impact that changed practice has on valued outcomes' (Timperley, 2008, p. 8). The same applies to teacher cooperative activities, which have been found to be effective only when geared towards solving an educational problem: 'As an intervention on its own, a

collegial community will often end up merely entrenching existing practice and the assumptions on which it is based' (p. 19; see also Vescio, Ross, & Adams, 2008).

Timperley et al. (2007) also found that the funding model employed for CPD had no bearing on its successfulness. The effectiveness of aligning teachers' needs to national and professional standards, research evidence, school operations and student outcomes was reported by Blank & de las Alas (2009), and teacher profiles may well be one way to reach this objective.

Even though evidence on what makes quality CPD is growing (along Timperley's work, see Adey, Hewitt, Hewitt, & Landau, 2004), it is harder to translate this knowledge into actual programmes. Interventions that have been shown to be effective in more than one circumstance exist (Allen, Hafen, Gregory, Mikami, & Pianta, 2015), but some of their mechanisms (e.g., their "dosage", Pianta et al., 2014) are still not well understood and in general these interventions are not widely adopted.

Even when robust study designs are employed, research on CPD is always affected by sampling and generalisability issues. Gersten et al. (2010) provide an insight into the practical difficulties that require overcoming in these kinds of studies, and one illustrative example is found in Heller et al. (2012), a randomized experiment which had to rely on a sample of volunteers. In general, few studies meet the strictest quality standards: Yoon et al. (2007) retained 9 out of the 1,300 pre-screened studies; Blank & de las Alas (2009) 16 out of 416; and Gersten et al. (2014) 5 out of 910.

Moreover, while there is unequivocal evidence that some forms of CPD improve teaching and learning (Joyce & Showers, 2002), there is one non-negligible aspect of the OECD advice that might prevent full policy implementation: including CPD in teacher evaluation. Despite big advancements, there is no consensus on the best approaches to evaluating teachers (Coe, Aloisi, Higgins, & Elliot Major, 2014; Darling-hammond, 2015; Hallinger, Heck, & Murphy, 2014; Raudenbush, 2015). Much of the debate revolves around finding the right balance between the formative and summative components of teacher assessment (Popham, 1988) and whether it is effectively possible (Master, 2014) or necessary (Hinchey, 2010) to separate them. The problem is that, at the bottom, teacher appraisal has to overcome not only methodological, but also political challenges 'dealing with issues of trust, authority, and knowing who is in charge of the information generated' (Coe et al., 2014, p. 26).

All things considered, perhaps the most interesting aspect about CPD is how little the OECD seems to be able to capitalise on an area in which almost all its recommendations are backed by independent evidence.

CPD was only briefly mentioned in the most recent TALIS reports (OECD, 2014e, 2014i) and recommendations focused only on the importance of mentoring and the lack of

opportunities and incentives. Of course, thematic reports have been produced (OECD, 2005c, 2011g), but compare the somewhat underwhelming coverage of CPD with the treatment reserved for an issue such as performance-related pay for teachers. When it comes to teacher pay, policy mechanisms were extensively and deeply analysed (Springer & Balch, 2009) and dedicated policy briefs (OECD, 2012b) sanctioned its salience, even though its effectiveness as a policy option is much more disputed than that of CPD. This may be a coincidence, and yet it feels as if, in this case, the importance of teachers' pay was "inflated" to the detriment of CPD because of its greater harmony with the economic slant (and expertise) of the OECD.

Autonomy and accountability

In PISA 2003, the OECD had reported that associations between autonomy and performance were generally weak, but also that autonomy in 'in certain aspects of school management' (OECD, 2004b, p. 236)—and especially in budget allocation—appeared to have some positive relationship with achievement. Similar claims were made in 2006. Autonomy in budgeting captured the greatest policy interest, although autonomy in curricular decisions began to emerge as a policy option (OECD, 2007b).

With regard to accountability, until PISA 2006, the OECD had measured some variables connected to accountability policies, including the existence of centralised external exit examinations (CEEEs) and whether schools' data were posted publicly (a practice often derogatively called "league tables" in educational literature). In 2006, both policies had a statistically significant association with student performance, although only the relationship with public posting of achievement data remained significant after accounting for demographic and socio-economic factors (OECD, 2007b). Although the section on accountability came right before that on autonomy in the international report, no links between the two were established.

Things changed in 2009. On the one hand, contrary to the policy preferences until then, a 'clear relationship' was highlighted 'between the degree of curricular autonomy a school system offers its schools and the system's performance' (OECD, 2010f, p. 41), whereas the viability of autonomy in budget allocation was suddenly questioned. On the other hand, a connection between autonomy and accountability was established. What had happened?

In those years, a series of secondary analyses revolving around the work of econometrician Ludger Wößmann (Fuchs & Wößmann, 2007; Wößmann et al., 2007; Wößmann, 2005) had introduced an interaction between autonomy and CEEEs which, 'in light of economic theory', was 'intuitively appealing' (Hanushek & Wößmann, 2011, p. 148). The argument goes as follows: in general, autonomy is negatively associated with outcomes, but the CEEEs make the association positive by 'introducing accountability and transparency and preventing opportunistic behavior in decentralized decision-making' (Fuchs & Wößmann, 2007, p. 452).

This is not a new idea: having a decentralised system that is controlled through incentives and sanctions is one of the staple characteristics of new public management (Dunleavy & Hood, 1994; Hood, 1991), but it is argued here that Wößmann's results allowed the OECD to establish a theoretical connection between autonomy and accountability.

Wößmann is the director of the Ifo Center for the Economics of Education (http://www.cesifo-group.de) and a senior expert at the European Expert Network on Economics of Education (EENEE, http://www.eenee.de), a think tank sponsored by the Directorate General for Education and Culture of the European Commission and coordinated by the very same Ifo Center. Andreas Schleicher also collaborates with the EENEE as an external advisor. Schleicher's collaboration 'gives the network a major possibility to co-ordinate its work with the crucial work at the OECD, to disseminate its research to the OECD and to receive feedback from the leading institution on internationally comparative measures of characteristics of education systems' (EENEE, 2015). It is plausible that Schleicher got to know Wößmann's work on CEEEs through the EENEE and "imported" it to the OECD. In fact, Wößmann's work started being referenced in international reports from PISA 2009.

The OECD did not simply reproduce the main arguments of this strand of econometric research; it integrated them into its own theory—which was still under development—of how autonomy interacts with accountability. For instance, while Wößmann (2005) noted that all autonomy measures (including curricular autonomy) are negatively related to PISA scores unless CEEEs exist in the system, the OECD continued to argue in 2009 and 2012 that curricular autonomy was a positive feature of education systems, and at most it mentioned *en passant* that this positive association 'can play out differently within countries and economies' (OECD, 2010f, p. 41, 2013g, p. 52)—a very disputable euphemism to say that there was also evidence of negative or no correlations.

The OECD also conceptually merged CEEEs and "league tables", whereas econometric literature keeps them separate (Boarini & Lüdemann, 2009). Specifically, in both 2009 and 2012 the OECD used a reference to Fuchs & Wößmann (2007) to introduce the core idea that 'schools with more autonomy tend to perform better than schools with less autonomy when they are part of school systems with more accountability arrangements' (OECD, 2013g, p. 18). By 'accountability arrangements' the OECD means the public posting of achievement data and the existence of standardised subject policies within schools, such as 'a school curriculum with shared instructional materials accompanied by staff development and training' (p. 53).

Fuchs & Wößmann (2007), however, used only CEEEs as a predictor variable and never mentioned public posting of achievement data in their paper. In 2009 the conceptual leap from econometric literature to the "league tables" was immediate, whereas in 2012 it was mediated

by the statement that CEEEs were ‘considered here as one form of accountability’ (p. 51). To add to the complexity, it should be noted that when the OECD links autonomy and accountability it is actually referring to two interactions: one is between public posting of achievement data and autonomy in resources, whereas the other between standardised subject policies and autonomy in curriculum and assessment.

After this excursus, one may wonder what other evidence exists about the effectiveness of autonomy and/or accountability arrangements. Re-analyses of PISA data reported mixed results. Boarini & Lüdemann (2009) found, using PISA 2006, a significant and sizeable positive relationship between CEEEs and performance, a smaller positive coefficient for public posting of achievement data, but no observable “effects” of autonomy—or autonomy interacting with accountability—on PISA scores. Using the same dataset, Hindriks et al. (2010) argued that within Flanders more autonomous schools tended to have better results, but the link with accountability had to be based on the presence of a system of school inspections because ‘there are no central examinations’ in the region (p. 26). Hanushek, Link, & Wößmann (2013) combined data from PISA 2000–2009 and argued that ‘the effect of *introducing* autonomy is more positive in countries that hold the system accountable by central exit exams’ (p. 226, emphasis added)—but they did not study the impact of “league tables”.

Scheerens et al. (2015) used multilevel structural equation modelling on the PISA 2009 data and suggested that presumed relationships between autonomy, accountability and performance were spurious and disappeared once school SES was introduced in the model. Benton (2014) demonstrated that, at least in the case of PISA 2009, the positive interaction between autonomy and accountability at the cross-country level was a statistical artefact caused by data aggregation, which disappeared when carrying out separate regression analyses for different countries. Outside the PISA dataset, however, Loeb & Strunk (2007) found that, in the US, autonomy over resources positively interacted with school accountability.

Even the literature focusing purely on CEEEs is divided. Wößmann continues to argue for their usefulness as a predictor variable in a range of settings (including labour market analyses, see Schwerdt & Wößmann, 2015), but Jürges et al. (2012) used a longitudinal extension to PISA 2003 and showed that CEEEs only improve curriculum-based knowledge and not overall literacy. Helbach (2012) merged the PISA 2000 and 2009 datasets and argued that student assessment through standardized tests (a proxy for CEEEs) has a positive association with country performance when interacting with other forms of school accountability, but not on its own. A research synthesis from the US reported no or negative effects of CEEEs on a range of performance and equity outcomes (Holme, Richards, Jimerson, & Cohen, 2010).

Finally, some of the bases upon which the reinforcing effects of autonomy and accountability are predicated can be questioned: will not ‘holding schools accountable through

national student testing [… practically reduce] their autonomy over curriculum' (Jensen, Weidmann, & Farmer, 2013, p. 37) because of a backwash effect? On a methodological level, to what extent are principal self-reports of school autonomy or accountability (used in PISA and other LSAs) a valid measure of the "real" variables (Brauckmann & Schwarz, 2014; Negrea & Duse, 2014)?

Student performance

Curricular changes

> The OECD advises that countries should introduce curricula that are 'better aligned with […] 21st century skills' (OECD, 2014g, p. 253).

Reviewing curricular priorities is an interesting recommendation, because on the one hand the OECD claims that PISA focuses 'on young people's ability to use their knowledge and skills to meet real-life challenges, rather than on the extent to which they have mastered a specific school curriculum' (OECD, 2001a, p. 14). On the other hand, 'this orientation reflects a change in the goals and objectives of curricula themselves which are increasingly concerned with what students can do with what they learn at school, and not merely with whether they have learned it' (*ibid*.).

The basic assumption is that traditional national curricula are self-centred and disconnected from what students would really need in their future life. "21st century" curricula, in contrast, share with the PISA literacies an applied and real-life problem-solving slant, and can therefore be tested by PISA. Thus, it is not the case that PISA does not assess mastery of 'a specific school curriculum' (OECD, 2001a, p. 14), but rather that it assesses mastery of *that* specific but implicit curriculum that has been agreed by the PGB from one cycle to the next. This is why the OECD felt entitled to praise a Canadian region for its new '21st century' curriculum (OECD, 2014f, p. 119): having a national curriculum is not an issue, the issue is having a curriculum that is different from the "PISA curriculum". The question, then, is what does the PISA curriculum comprise? What is the nature of the so-called "21st century skills"?

Schleicher (2007) acknowledged that 'there is no overarching agreement on what fundamental competencies 15-year-olds should possess' (p. 352), but it is clear that, according to the OECD, 21st century skills must include the PISA literacies. After participating in PISA, many countries attempted to align their national curricula to it. This phenomenon was highlighted in Chapter 1, but a deeper analysis can be made, taking scientific literacy as an example. Some policy responses included reforms in Denmark (Dolin & Krogh, 2010) and the Netherlands (Eijkelhof, Kordes, & Savelsbergh, 2013); the US *Framework for K-12 Science Education*, whose core ideas 'are consistent with frameworks for national and international assessments' like PISA (National Research Council, 2012, p. 141); or the aptly

named “Twenty First Century Science”, ‘a set of GCSE science courses giving all 14-16-year-olds a worthwhile and inspiring experience of science’ (Nuffield Foundation, 2015).

Dillon (2009) argued that “scientific literacy” is a slippery concept that can be fit for multiple purposes, which may explain why it found its way in many curricular reforms regardless of their actual correspondence to what the OECD intends by the phrase. Other authors have questioned the accuracy and appropriateness of the construct “science” and the extent to which PISA manages to assess it (Kind, 2013; Lau, 2009; Le Hebel, Montpied, & Tiberghien, 2014; Serder & Jakobsson, 2014), though a discussion is outside the scope of this section.

While many national curricula are being aligned with the PISA literacies, it is unclear whether what PISA actually tests is a washed down version of national curricula. Lau (2009) critiques the lack of information from the OECD about the extent to which PISA was ‘informed, but not constrained by the common denominator of national curricula’ (OECD, 2006, p. 7). He turns one of the purported strengths of PISA (the fact that country scores are not affected by the curricular relevance of the PISA items, see McGaw, 2008) into a possible liability: if the OECD always tries to minimise the variability of the items’ curricular relevance, how can it be claimed that PISA is not measuring the common denominator of national curricula (Lau, 2009)?

In spite of the influence of the PISA framework on country curricula, and of country curricula on PISA items, it is not known whether curricular alignment increases country performance—let alone whether employers will actually value the skill sets promoted by the OECD and others (like The Partnership for 21st Century Skills, www.p21.org), which is the assumption at the basis of this paradigm shift. As Ladwig (2010) reminds us, calls for ‘nonacademic’ outcomes outside the mastery of traditional subjects are not new in education. Skill-based curricula always bring back questions of ‘just how much […] we really want schools to do’ (pp. 113–114) and about the wider purposes of education. These kinds of curricula would have to be evaluated from the extent to which they are successful ‘both in producing schooling outcomes and in the intended subsequent social effect’ (*ibid*.).

It is too early to know what social effects could be linked to high PISA performance, and evidence of the predictive validity of PISA exists but is limited. Studies with a longitudinal component in Canada (Knighton & Bussière, 2006) and Denmark found that higher PISA scores were associated with increased odds of completing secondary education. In the International Adult Literacy Study, employment status and earnings were more strongly correlated with PISA literacies than with educational qualifications (Schleicher, 2007). Fischbach et al. (2013) showed that, in Luxembourg, higher PISA scores were associated with higher school grades and lower probability of repeating a class, and reported that a link

between PISA scores and grade repetition or school completion was also found in Germany and Australia. Jakubowski (2013) highlighted that, in Australia, PISA items purporting to measure higher thinking skills were related to future student qualifications, but he had to concede that it was too early to know what the relationship between PISA and 'the most desirable outcomes' (p. 25), such as labour market outcomes, could be.

The strongest critique to this research strand is that it only shows that the best students also score high in PISA. If anything, this is evidence that PISA measures some general academic ability, rather than 21st century skills. Furthermore, none of these studies related student attainment to curricular *changes*, which means that it is not known whether following the EPA has any effect on student outcomes. Note that if any such effects were observed, they would still be difficult to interpret: they could be read both as a positive outcome (students are becoming more "literate") or as a negative one (you can teach to PISA).

This issue is tackled in more detail in the next chapter, but in general, the problem is that, without an external frame of reference, one cannot know whether PISA can be affected by score inflation in a similar way as high-stakes tests are. Brunner et al. (2007) used a quasi-experimental design to show that PISA 2003 scores could be improved by 9—24 by pre-testing students and coaching them for two weeks before retaking the assessment. More recently, Schleicher (2013f) fiercely rebutted claims that the success of Shanghai was due to teaching to the test, arguing that the PISA 2012 technical report (OECD, 2014h) found no sign of malpractice.

A final point that is worth drawing attention to concerns another part of the OECD recommendation, namely the suggestion that curricula should be 'better aligned with students' interests' (OECD, 2014g, p. 253). As an example, the OECD says:

> 'Japanese students are interested in inquiry-based learning, whereas science teaching at the upper secondary level does not cater to that interest. This finding could contribute to an understanding of why Japanese students in PISA show relatively low levels of positive attitudes toward science.' (OECD, 2013c, p. 172)

However, Ogura & Goto (2013) reported that Japanese students in grades 10 and 12 simply value scientific tracks less than other options, which may explain their low levels of positive attitudes towards science. Perhaps the issue in Japanese education is not lack of inquiry-based learning but lack of interest in a scientific career. Would the OECD still claim that curricula should be better aligned with students' interests when students' interests do not involve one of the PISA literacies? It is likely that the actual OECD advice is not, as written, that governments should adapt the literacies to what students might consider of interest, but rather that they should try to convince students that the OECD literacies are interesting.

Besides, aligning curricula to perceived students' interests to make them more relevant is a difficult endeavour which is currently not supported by solid theoretical underpinnings. In

some Australian regions, similar changes made student assessment less valid and tended to exacerbate the differences between more and less academically-minded students (Fenwick, 2011). When listening to students' interests becomes personalised learning, there is a risk that education is turned into a market in which only the savviest customers will be able to thrive (Prain et al., 2013).

Conclusions to Chapter 3

This chapter has reconstructed the OECD education policy advice and has briefly evaluated its internal coherence, consistency over time and soundness.

The coherence of the EPA

The EPA is difficult to frame accurately. Whilst made of explicit recommendations, as a whole the policy advice emerges from OECD publications without ever having been clearly summarised and organised as was done for this thesis. It feels as if it were the result of several contributing voices over the years, from within and outside education, without a unique plan for country education. In fact, the Strategic Development Group argued over ten years ago that 'the key policy objectives that are driving the development of PISA' (OECD, 2003c, p. 10) needed to be conceptualised better, and that the policy levers that could help countries to achieve these objectives had not been explicitly identified.

There are a few generic core areas of interest that recur over time and across publications; issues related to disadvantage, lifelong learning, or teacher quality. Many other areas instead are given attention only intermittently. A certain "fuzziness" characterises the boundaries delimiting the educational topics and subtopics that are the object of the OECD recommendations (and recall the fuzziness of the very concept of "education policy"). Similarly varying is the modality with which the advice is given: the OECD practical arguments can involve some very explicit and specific recommendations, some implicit and general ones, and many other variations along the axes of explicitness and specificity. Overall, the OECD is a very prolific advisor, but the outcome is a 'patchwork of loosely connected topics [… that] have been strung together to provide a minimum semblance of coherence' that Henry et al. (2001, p. 52) poignantly recognised.

This author would argue that the problems with the coherence of the EPA are due to the fact that educational concepts (e.g., "accountability", "choice" or "leadership") are *symbols*, 'vehicles [… carrying] a load of "meanings"' (Turner, 1975, p. 152). Educational concepts appear transparent but they actually hide multiple interpretations; they originate from different fields (e.g., the econometric or educational effectiveness literature) and accordingly they may acquire different meanings.

Educational concepts and related indicators (other symbolic constructions) are supposed to be value-free, and this is an advantage for the OECD because 'the less obviously political in form symbols are, the more efficacious politically they prove to be' (Cohen, 1979, p. 87). Since symbols are 'multivocal, manipulable, and ambiguous' (Turner, 1975, p. 146), so are symbolic systems like the framework embodying the OECD educational understanding and policy stance. Symbolic systems 'confront us as lacking coherence, "with gaps, modifications, and inconsistencies" […] because of "the continual intrusion of pragmatic issues into the kinds of functions symbols are made to serve"' (Firth, 1973, p. 426; quoted in Turner, 1975, p. 146).

Turner (1975) argued that order in symbolic system comes 'from purpose, not from *connaissance*' (p. 147). Given the complex internal functioning of the OECD, its recurrent identity crisis (Carroll & Kellow, 2011; Marcussen & Trondal, 2011) and the novelty of education as a primary branch, it is not surprising that the EPA is "disordered". Nevertheless, it is a stain on the mantle of scientific rationality that the OECD supposedly wears (M. N. Barnett & Finnemore, 1999).

The consistency of the EPA

The analysis showed that the EPA is remarkably stable over time, but the extent to which this was the result of happenstance as opposed to a purposeful decision could not be investigated within this research. Nevertheless, some tentative interpretations can be advanced.

The priorities of the PGB may play a role when evidence is scarce. If priorities do not change and there is no strong evidence for or against a policy, then some presumptive inferences may be reproduced from cycle to cycle until they are disproved. Examples include recommendations on school funding, class size or teacher qualifications, which continue to be included in PISA reports in spite of a limited ability of their indicators to provide valuable information.

There are claims that teacher training and qualifications are 'most important for educational policy, because there is a growing need to recruit teachers from non-traditional backgrounds' (Klieme & Kuger, 2014, p. 28). However, they have also been found to be unrelated to teacher quality or effectiveness (Hanushek & Rivkin, 2006, 2012). Class size (Leithwood & Jantzi, 2009) and funding (Gibbons & Mcnally, 2013) may or may not be effective depending on the context and form of implementation, but they are still measured by PISA.

Perhaps it is because of their political (rather than educational) value; perhaps they are simply part of a heritage that has thus far gone unquestioned. An interviewee from one of the case studies introduced later on in this document mentioned that the OECD does not draw only from empirical research, but also from educational "fashions" and values. For instance,

the rationale for including class size as a measure is that it has 'always been included in the School Questionnaire' (Klieme & Kuger, 2014, p. 33). The combination of low informative power and a constant interest would provide little incentive to vary the recommendations, which would explain why some of them are reproduced verbatim across publications and over time.

Another (and complementary) possibility is that, in areas that have attracted less interest from the PGB, the Secretariat may have enjoyed more autonomy and found more openings to introduce and reiterate some key concepts shared by the research community within the OECD. This is more an impression informed by the findings than a research hypothesis, and besides, there is evidence refuting it. For instance, competition and choice have never been top PGB priorities for PISA, but rather than disinterest this may reflect the fact that they are highly contested options. Issues of schools competing for their students created several tensions in the 1980s between egalitarians and neo-conservatives (Henry et al., 2001). At present, international reports carefully balance explanations for either view, although the shortcomings of school competition are generally given more room.

The role of PISA in informing the EPA

The analysis found inconsistent presentation of statistical information, cluttered graphs and some unidirectional reading of correlational data in international reports. Moreover, success stories were sometimes used to promote the usefulness of PISA as a source of information, when in fact PISA or the EPA played a limited role in informing the policies of exemplar countries. Altogether, there were indications that the EPA might pre-date PISA evidence, instead of being informed by it.

This is consistent with Pedró's (2012) claim that policy advice from PISA is not evidence-based, but rather a 'leap in the dark'. Pedró (2012) observed that strong claims are often build upon 'feeble' bases (p. 157) qualified by a disclaimer. Sometimes, this practice is taken even further: Pedró (2012) brings the example of praise for public posting of achievement data in the 2006 report which was directly contradicted by the evidence in the same chapter. In sum, the OECD seems to be able to make, in official publications, the same instrumental use of PISA as countries do.

The stability of the EPA *over and above* PISA evidence may be necessary for constructing successful knowledge for policy, but it certainly testifies against the validity of PISA in producing data for policy. It does produces data, it is indeed used—or at least mentioned—in policymaking, but the 'bridge' (Pedró, 2012) between data and policy is sometimes created ad hoc. This is a definite consequence of the poor theoretical framework critiqued in the previous

section. In these circumstances, the EPA becomes little more than a policy narrative, a story with which to make sense of the unknown (Roe, 1994).

The soundness of the EPA

The analysis explored the soundness of the EPA in selected recommendations. The overarching theme was once again the lack of structure in the framework, which is captured by four interrelated issues.

Strong evidence not capitalised on

Pre-primary education is an example of a recommendation in which very strong scientific evidence is not matched by an (official) interest of the PGB. This variable was not measured in 2000 and 2006, and although it was part of the 2015 framework it was not assigned a high-priority status. It is unclear why this may be the case. Naturally, early childhood provisions are different in any country, but so are priority indicators such as leadership and school management (Klieme & Kuger, 2014). Moreover, 'every country in Europe has set up some form of publicly subsidised and accredited early childhood education and care for children below the age of compulsory schooling' (Eurydice, 2009a, p. 75). Perhaps the PGB is aware of the importance of pre-primary education, but perceived difficulties in policy action may result in this component being assigned lower priority: countries might know that with or without OECD backing, they can do little to change the existing situation.

Other examples are stratification policies such as tracking and retention, where the consensus is that they have a negative impact on equity and no impact or negative impact on achievement. The OECD acknowledges and insists on this point (for instance through PISA in Focus, see OECD, 2011i, 2014a, 2014b), but the quality and strength of the EPA do not seem to match those of the evidence. When practical advice is given—such as to provide 'extra teaching time for students who fall behind' (OECD, 2014a, p. 4) or ensure transferability between the academic and vocational pathway (see Table 3.2)—little research exists to confirm or contradict such statements. A peculiar case is student transferring as a disciplinary action, an example of ineffective policy used by the OECD to advocate for more school autonomy with no clear theorisation linking the two.

A third example is CPD, a practice finding great support in the literature (provided it has certain qualities) and correspondingly limited attention from the OECD, especially in comparison with other policies (teacher pay) whose evidence base is far more disputed. It was argued that this may be due to the prevalence of the econometric experts within the organisation, who may be less aware of topics traditionally belonging to the education field.

Theorisation becoming better over time

At the beginning of PISA, educational areas such as gender gap lacked theoretical support. The conceptualisation of gender gap was confusing because the OECD problematised it in terms of both higher education choices and secondary education achievement. The magnitude of the gap in secondary education is uncertain. The recommendation about making certain pathways more appealing to the underrepresented gender was appropriate, but insufficient, since the source of the problem lies in a combination of individual and environmental factors.

More recently, the OECD substantively improved its theoretical framing of the issue. As a result, it may be able to give more targeted advice in the future. Some positive signals in this direction come from *The ABC of Gender Equality in Education* report (OECD, 2015i), whose policy recommendations were perhaps not very cohesive, but they were focused, clear and consistent with research findings.

EPA still under development

School competition and choice have been contested topics for many years, but vouchers have entered the OECD radar only recently. They were practically absent from international reports until 2009, when they were described as a possible school choice (OECD, 2010f). In 2012 a stronger link between vouchers, choice and inequality was made, but it was not accompanied by any specific policy recommendation (OECD, 2013g). In the *Education GPS*, however, they are included as a policy option which facilitates (voluntary) student transfer, whereas *PISA in Focus* suggests that they can work as socio-economic equalisers when they target specific disadvantaged groups (OECD, 2012a). The reference is to the 'progressive' voucher schemes (Musset, 2012; OECD, 2012c), but a clear distinction between progressive and non-progressive schemes is drawn only in some publications.

Overall, it seems as if the OECD were beginning to introduce vouchers as a viable policy option that still needs being framed appropriately, especially since they belong to a category (competition and choice) which is contested in the literature and by some member countries.

A similar argument can be made for school autonomy. Initially, findings on different types of autonomy (resources, curriculum, assessment, staffing) were presented together as if "autonomy" only meant one thing, although budget allocation was given somewhat more prominence. Progressively, data on autonomy were disaggregated and curricular autonomy was placed under the spotlight. At the same time, a connection was established between autonomy and accountability, perhaps drawing from econometric research but replacing the role of CEEEs with public posting of achievement data as a measure of accountability (see for example OECD, 2011f).

An evolution of the EPA is welcome; in fact, it is exactly what should happen when the evidence changes. However, the impression in this instance is that a strong EPA was absent until the opportunity came to connect autonomy and accountability within a New Public Management framework. This may have the double effect of clarifying but also crystallising the EPA.

PISA for PISA's sake

When the OECD suggests that country curricula should be more aligned with 21st century skills it is implicitly conceding that it is not true that PISA is curriculum-free, only that it follows its own. Some would frame the undefined "PISA curriculum" within the neo-liberal agenda, but for the purpose of this study it is sufficient to recognise that it includes the core literacies (Reading, Mathematics and Science) and other skills (Problem-solving, Finance) that are less frequently assessed. Unfortunately, it is not possible to determine the extent to which the PISA curriculum is not a watered-down and amended version of national curricula without studying each country's curriculum.

However, it can be argued that at present there is no evidence that aligning national curricula to vaguely-defined 21st century skills would improve country performance in PISA—let alone have an impact on labour market outcomes as "promised" by the OECD. The fact that different studies found PISA scores to be predictive of a lower risk of school dropout may simply indicate that PISA captures general academic ability rather than skills for life. Finally, it was observed that when the OECD suggests that curricula should reflect the students' interest, in reality it is restating the importance of PISA by arguing that governments should motivated students to embrace its curriculum.

The next chapter

Some authors (Baird et al., 2011; Pedró, 2012) have noted that the policy effects of PISA—and the value of the EPA—can only be appreciated through trend analyses and triangulation of different evidence sources. Chapter 4 provides a cross-national measure of country alignment with PISA recommendations (the EPA Index) and different approaches to trend analysis to estimate, albeit indirectly, the effectiveness of the EPA. As part of the evaluation, the validity and the limitations of PISA as an instrument to measure change are also addressed. Cross-national evidence from Chapter 4 is complemented by evidence from two case studies in Chapters 5 and 6.

Chapter 4: A cross-country analysis of the use and impact of the EPA

This chapter is articulated in the following parts:

- A cross-country analysis of trends in performance and equity.
- The relationship between specific policy outputs and PISA outcomes.
- An exploration of country alignment with the EPA.
- A summary of the limitations of PISA.

There is evidence of policy alignment with the EPA in some countries and cross-country trends in performance have been on the rise, though only moderately. However, there is less evidence that countries are implementing OECD recommendations in terms of equity, though there are indications that doing so might be associated with changes in performance. While the reliability of PISA is high, it is argued that score differences below 20 points across countries or over time may often be a statistical artefact, and should not be overly-emphasised in policy reports. Finally, a significant correlation was found between country expansion of pre-primary education and increases in PISA scores ten years later. This adds to the evidence on the positive effects of early childhood education widely reported in the literature, and makes one question the decision of the PGB not to consider it a policy priority, as was noted in Chapter 3.

Introduction

In this chapter, large-scale evidence on the use of EPA for national policymaking and its plausible impact on country outcomes is sought by comparing original data analyses[11] with findings from other studies. The chapter proceeds in three steps that use increasingly targeted

[11] Except where indicated, all analyses in this chapter were original work, carried out within the statistical environment R (R Core Team, 2015) using the "nlme" package (version 3.1-117, Pinheiro et al., 2015) and the "ggplot2" package for charts (Wickham, 2009). The analyses used the PISA datasets that are freely available from the OECD website (https://www.oecd.org/pisa/pisaproducts/, accessed 16 April 2016). The equity measures used for the analyses in the "First step: trends in performance and equity" section, below, were not available for all country-years. Therefore, a new dataset comprising the Strength and the Slope of the socio-economic gradient (see equity section) for all countries and all years was produced using the statistical software IBM® SPSS® Statistics 22 and the methodology available in the PISA *Data Analysis Manual* (OECD, 2009d). Data on country GDP and gross enrolment ratio at pre-primary education were taken from The World Bank and UNESCO databases. For more information on the multilevel modelling approach used in this chapter, refer to Appendix 5 and 6.

data. The first step models educational outcomes in performance and equity; the second relates outputs to outcomes; and the third links policy alignment with outcomes.

The first step is to analyse trends in performance and equity since PISA started in 2000. If the EPA is adopted, implemented and effective, then some increases in performance or equity should be observed. In this chapter, score and inequality changes over time are modelled using correlational analysis and multilevel growth models. Particular emphasis is given to interpreting the linearity of trends, the clustering of OECD and partner countries, and the stability of the within-country variance compared to the changes in the between-country variance.

The second step contrasts the relationships of three policy outputs (general economic productivity, equity levels and access to pre-primary education) with PISA outcomes. The rationale is that any detectable "effect" on PISA scores should be larger for the factor that is more malleable by specific education policies (access to pre-primary education), smaller for the "generic" educational factor (equitability) and even smaller for the non-educational one (productivity). By comparing the effects of the three indicators, the second step attempts to set an empirical limit on the extent to which education reforms could be expected to produce measurable changes in country outcomes.

The third step complements secondary PISA data analyses with information collected through expert questionnaires, in order to relate country levels and changes in achievement and equity to their degree of alignment with OECD recommendations. The questionnaires return both an overall rating of country alignment with the EPA and a breakdown for a small sample of countries.

Finally, the chapter considers validity and reliability issues in relation to PISA, in particular issues affecting the comparability of scores between countries and over time. The OECD response to some of the PISA shortcomings is discussed, and conclusions are drawn.

First step: trends in performance and equity

Trends in performance

Figure 4.1 and Figure 4.2 map trends in Reading between 2000 and 2012, and Mathematics between 2003 and 2012. Scores in the Reading domain are comparable across five administrations and have the highest variability, yet trends look very linear. Most changes are small, but there is evidence of some large score increases, especially among non-OECD countries. The red and blue lines show the overall gap between OECD and partner countries. In Reading, the dots on occasion 0 (the first comparable PISA cycle) appear more spread out than they do on occasion 4 (PISA 2012), which suggests that results may be converging with

time. In Mathematics, large changes are rarer and the spread seems more constant than for Reading, but the difference between the OECD cluster and other countries is still visible. Science trends (not displayed) are very similar to those in Mathematics.

Figure 4.1: Trends in PISA Reading scores by measurement occasion. A value of 0 corresponds to the year 2000.

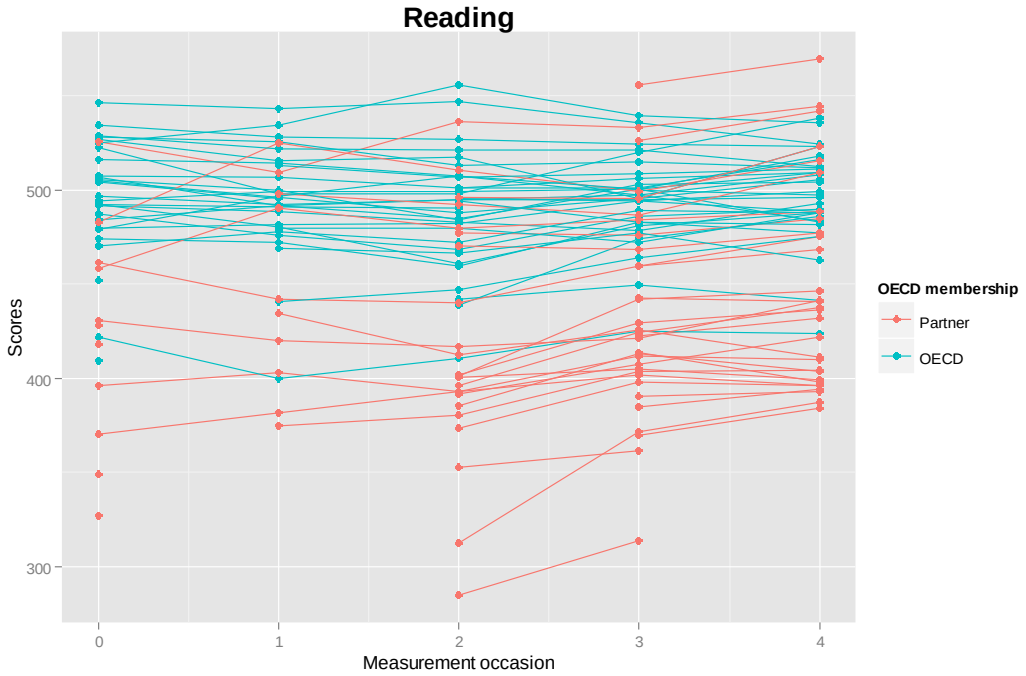
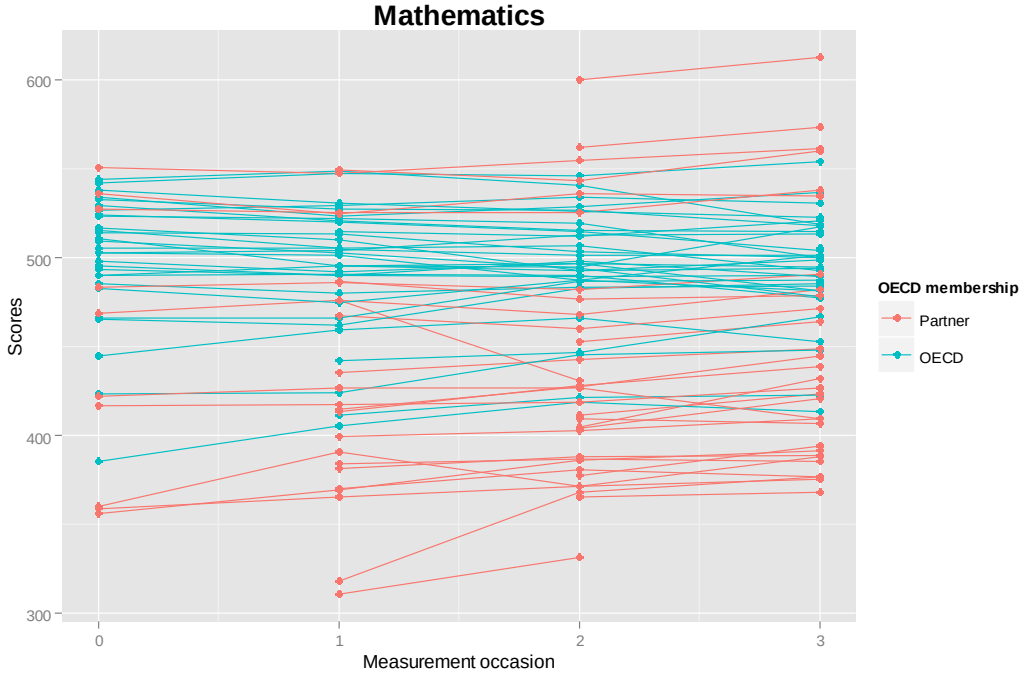


Figure 4.2: Trends in PISA Mathematics scores by measurement occasion. A value of 0 corresponds to the year 2003.



Sources: own elaborations using PISA data

The impressions of linearity, of differences between OECD and partner countries and of a slight decrease in spread over time are supported by correlational and multilevel analyses, which also provide some indications of the magnitude of score changes over time.

Correlational analysis: change between two time points

*Table 4.1: correlation between country mean PISA scores, using the full (**D_0**,* n = *66) and reduced (**D_1**,* n = *30) datasets*

			FULL DATASET (D_0)											
			READING 2000–2012					*MATHS 2003–2012*				*SCIENCE 06–12*		
		Year	*00*	*03*	*06*	*09*	*12*	*03*	*06*	*09*	*12*	*06*	*09*	*12*
REDUCED DATASET (D_1)	*READING*	*00*	1	.94	.91	.95	.93	.92	.92	.93	.89	.93	.94	.92
		03	.94	1	.97	.96	.92	.95	.94	.94	.90	.95	.95	.92
		06	.92	.97	1	.97	.94	.94	.96	.94	.91	.96	.96	.94
		09	.94	.96	.96	1	.97	.93	.95	.96	.93	.95	.98	.96
		12	.91	.92	.94	.95	1	.92	.94	.96	.96	.94	.97	.98
	MATHS	*03*	.92	.94	.93	.94	.94	1	.99	.98	.95	.97	.96	.94
		06	.91	.94	.95	.95	.94	.99	1	.98	.96	.98	.97	.95
		09	.90	.93	.94	.96	.95	.97	.97	1	.98	.96	.98	.97
		12	.84	.88	.91	.92	.97	.95	.95	.97	1	.94	.96	.97
	SCIE	*06*	.92	.94	.94	.95	.94	.97	.97	.96	.92	1	.98	.97
		09	.92	.94	.95	.97	.97	.96	.96	.98	.95	.98	1	.99
		12	.88	.91	.92	.93	.98	.94	.93	.95	.97	.96	.98	1

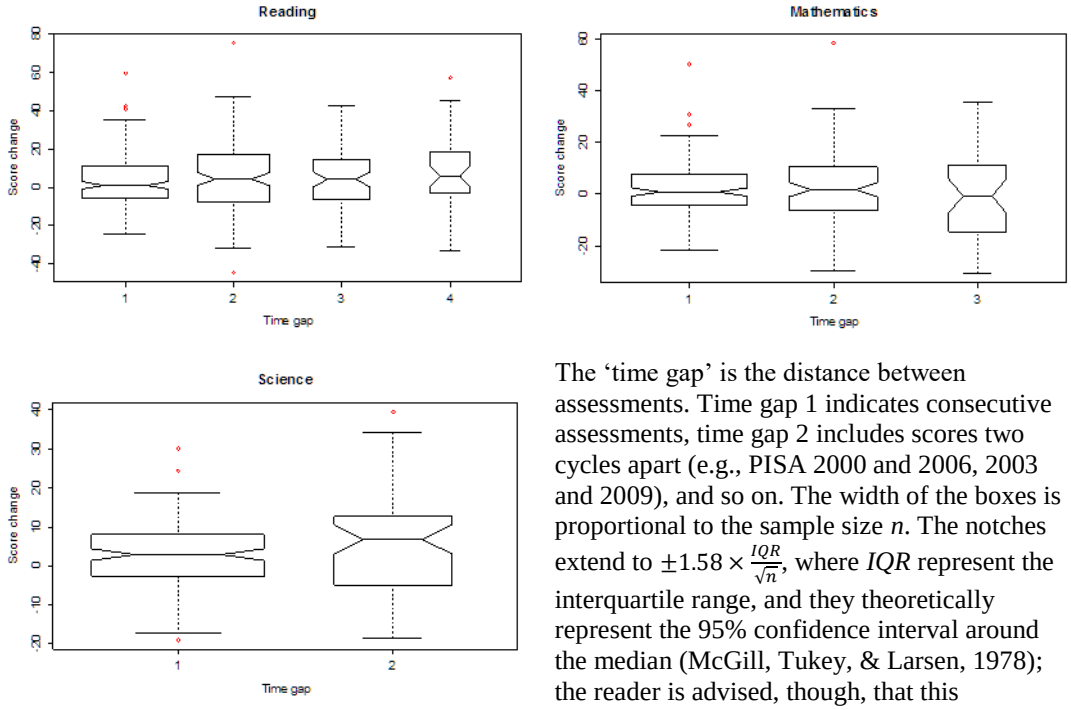
Source: own elaboration using PISA data.

Table 4.1 summarises the Pearson's correlation coefficient *r* between country mean scores in Reading, Mathematics and Science from PISA 2000 to PISA 2012. The Reading scales are comparable between 2000 and 2012; the Mathematics scales between 2003 and 2012; and the Science scales between 2006 and 2012 (OECD, 2009d). The correlations in the upper-right part of the table were computed using a dataset of all countries with at least two observations (**D_0**, $n = 66$, Appendix 3, Table 4.A) with pairwise deletion of missing data, which means that the subsets being correlated between each pair of assessment cycles included different countries. To account for this, the bottom-left side of the table shows correlations from a reduced dataset (**D_1**, $n = 30$) comprising only complete cases—i.e., only those 30 countries that took part in all PISA assessments.

Correlations using either dataset are highly comparable and very strong, ranging from .89 to .99 in **D_0** and from .84 to .99 in **D_1**. This suggests that countries experience very little variation in scores within and across domains over time. These correlations do not depend on the standard errors of the mean country scores (analysis not shown). Within subjects, correlations decrease with some regularity, which is what one would expect if scores were changing in a linear fashion. Between subjects, correlations in Mathematics and Science are very similar, suggesting that scores in one domain "explain" with a high level of accuracy the scores in the other. To have an idea of the magnitude of the score variability plotted in Figure

4.1 and Figure 4.2, the distribution of score changes between different PISA cycles is shown in Figure 4.3.

Figure 4.3: Distribution of score changes by time gap and domain



The 'time gap' is the distance between assessments. Time gap 1 indicates consecutive assessments, time gap 2 includes scores two cycles apart (e.g., PISA 2000 and 2006, 2003 and 2009), and so on. The width of the boxes is proportional to the sample size *n*. The notches extend to $\pm 1.58 \times \frac{IQR}{\sqrt{n}}$, where *IQR* represent the interquartile range, and they theoretically represent the 95% confidence interval around the median (McGill, Tukey, & Larsen, 1978); the reader is advised, though, that this calculation is only an approximation and that the images are only indicative. Outliers are plotted in red.

Source: own elaboration using PISA data

The boxplots show that the magnitude and variability of score changes tend to increase the further apart two assessments, confirming that score changes, especially negative ones, tend to be incremental rather than sudden. Large improvements (or decreases) from one cycle to the next are possible but unlikely, and in general, most country score changes are well below ±20 points on the scale with a SD of 100 (Lenkeit & Caro, 2014, calculate that 95% of them are between -13.8 and +17.8 points).

Some outlying score changes are shown in Table 4.2. The fact that extreme score changes are observed mostly in Reading and among non-OECD countries raises a question: have these countries implemented particularly effective policies targeting reading literacy, or are the data for Reading less reliable?

Table 4.2: a list of country score changes above ± 40 points

CHANGE	COUNTRY	OECD?	DOMAIN	PERIOD
-44.53	Argentina	No	Reading	2000–2006
40.09	Japan	Yes	Reading	2006–2012
40.99	Serbia	No	Reading	2006–2009
41.67	Romania	No	Reading	2006–2012

42.49	Liechtenstein	No	Reading	2000–2003
42.61	Peru	No	Reading	2000–2009
45.10	Serbia	No	Reading	2006–2012
45.11	Albania	No	Reading	2000–2012
47.13	Israel	Yes	Reading	2006–2012
50.16	Qatar	No	Maths	2006–2009
57.07	Peru	No	Reading	2000–2012
58.49	Qatar	No	Maths	2006–2012
59.502	Qatar	No	Reading	2006–2009
75.29	Qatar	No	Reading	2006–2012

Source: own elaboration using PISA data

A cursory examination of the PISA reliability coefficients provided by the OECD (Appendix 3, Table 4.B) shows that the Reading scale is as reliable as the Mathematics and Science scales, but a more thorough discussion about the robustness of PISA across changing times and conditions, which is presented at the end of this chapter, returns a more complex picture. Moreover, the case of Qatar suggests that the actual impact of (EPA-aligned) policies may be lower than what the numbers in Table 4.2 imply.

Qatar began its "Education for a New Era" reform in 2001. The reform was designed by the RAND Corporation following international recommendations summarised in Guarino & Tanner (2012) as the 'principles of adequacy in funding, accountability, autonomy and equity' (p. 224). The major intervention was the creation of publicly-funded independent schools pairing autonomy in curriculum and staffing with increased accountability (Zellman et al., 2009). Four subjects were compulsory (Arabic, mathematics, science and English), but teachers and principals were otherwise given pedagogical and curricular freedom, as well as additional CPD opportunities. Independent schools were also granted more powers for hiring, paying and dismissing teachers. Accountability was established through evaluations, standardised assessments and school reports cards that would allow parents to make informed choices.

The first independent schools were established between 2004 and 2005, and a first evaluation by the RAND Corporation was completed in 2007. Independent schools had managed to increase the percentage of students approaching governmental standards in Arabic and English, though self-selection of highly motivated teachers and students in the new schools could not be excluded. The implementation of the reform was also hampered by systemic resistance and unintended consequences that made evaluators doubt whether any improvement could be sustained over the following years (Zellman et al., 2009).

Using national data from 2004–2006, Guarino & Tanner (2012) reported that 'many of the reform principles that reflected international recommendations were almost immediately transformed – largely in the direction of mimicking the pre-existing system' (p. 243), which

severely reduced the effectiveness of the intervention. 'Initially generous funding levels were reduced, […] provider autonomy was curbed and the population from which providers could be drawn […] was limited to individuals within the system' (*ibid.*). Accountability 'was enforced primarily via spending restrictions', whereas 'the parental choice mechanism […] was relatively weak' (p. 242). Gender parity was hardly achieved.

Despite the many shortcomings, Guarino & Tanner (2012) thought that the reform had planted some seeds for future achievement. Indeed, Qatar improved its performance by up to 75 points in PISA and by over 100 points between TIMSS 2007 and 2011 (Mullis, Martin, Foy, & Arora, 2012; Mullis et al., 2008). However, the performance gap between public (including independent) and private schools in favour of these latter remained the same (Cheema, 2015), the rate of improvement has been decreasing and gains between PISA 2009 and 2012 are explained almost completely by changes in the socio-economic and demographic composition of the student population (OECD, 2014g).

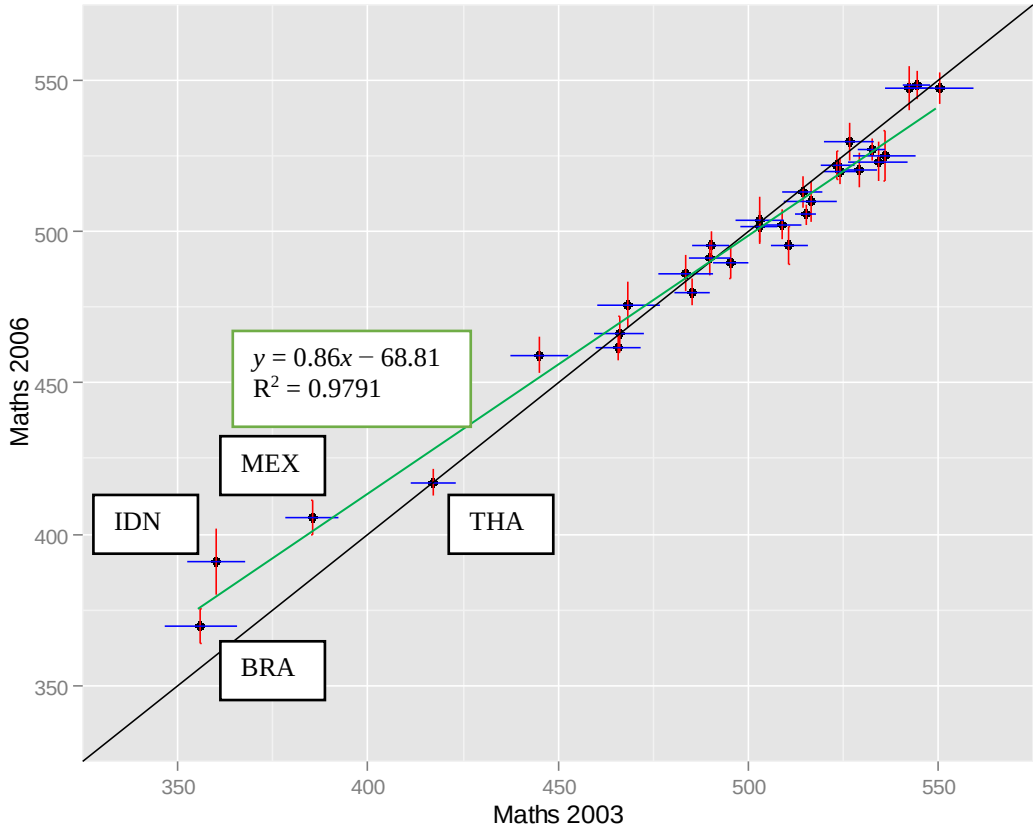
Furthermore, it is not unlikely that changes in the performance of immigrant students may be the main driver behind PISA trends, irrespective of any educational reforms. In Europe, 20% of the PISA 2009 Reading items were easier for immigrants, even though they are generally a disadvantaged minority (Araújo & Dinis da Costa, 2012). In Qatar, immigrant students are the majority, have higher motivation and consistently outperform natives in all PISA literacies, Reading in particular (Areepattamannil, Melkonian, & Khine, 2015; Areepattamannil, 2012; Cheema, 2014). Response patterns in the English version of PISA 2006 were closer to those of pupils in English-speaking countries than to those of Qatari students taking the Arabic version of the assessment (Grisay, De Jong, Gebhardt, Berezner, & Halleux-Monseur, 2007).

Overall, the "Education for a New Era" reform was originally in agreement with the EPA but several adjustments during its implementation rapidly brought it out of line. There is evidence of performance (but not equity) increases in LSAs, but the reform did not bridge the gap between public and private education and did not affect the behaviours and outcomes of Qatari students—its intended target population. It is hard to claim evidence of impact or that any impact was due to policy alignment with the EPA. In fact, an alternative narrative might state that once the Qatari leadership amended the education reform so that it was *less* aligned with the EPA, country performance improved.

Other questions about the relationship between PISA scores and the effectiveness of country policies are raised by Figure 4.4, which plots results from Mathematics 2003 and 2006. The correlation between the two sets of value is .99 in $\mathbf{D}_0$ and $\mathbf{D}_1$, the highest among all PISA correlations in Table 4.1. The plot shows the 95% confidence intervals around the

country scores in 2003 (horizontal blue bars) and in 2006 (vertical red bars), as well as the regression line (in green) and the principal axis (in black)[12].

Figure 4.4: The highest score correlation, Mathematics 2003–2006



Source: own elaboration using PISA data

Very few countries (such as Brazil, Indonesia, and Mexico) experienced statistically significant changes in Mathematics between these two PISA cycles. Is it reasonable to maintain that only these countries managed to introduce effective policies? And to what extent do changes between two time points reflect actual increases in student ability? Possible answers are explored below by considering the case of Indonesia.

Indonesia overhauled its education system in 2003. New legislation introduced competency-based curricula, a more stringent teacher certification system, as well as increased

[12] While the regression line is the line of best fit, the principal axis is not a summary of the data but a statement based on equality: it is the line that would be obtained if scores in 2006 were equal to those in 2003. The regression line is often a less-than-ideal method to calculate difference in performance between two time points (He & Tymms, 2013). Consider the case of Thailand: Thailand falls exactly on the principal axis, meaning that its scores in 2003 and 2006 were almost identical (and indeed, they were 416.978 and 417.073). This is not evident from the regression line. When comparing two time-points, knowing that Thailand progressed less than expected considering all data-points (an inference about space) is arguably less informative than knowing that it progressed less than expected considering its previous performance (an inference about time): the principal axis helps to visualise changes in scores from one PISA wave to the next.

autonomy in governance and pedagogy (Mullis, Martin, Minnich, et al., 2012; UNESCO-IBE, 2011a). This, along with improvements in Reading and a reduction in the percentage of low achievers, earned the country some praise from the OECD (2011c). Indeed, Reading scores increased by 31 points between 2000 and 2009 and Mathematics scores also increased by 31-point between 2003 and 2006 (the gap shown in Figure 4.4). Barrera-Osorio et al. (2011) focused on this gap and found that the greatest share of score variance between the two cycles was not explained by school and student characteristics or background. They concluded that 'Indonesia was able to better educate its students in 2006 than in 2003' (p. 18).

However, they did not consider the alternative explanation that the observed changes might have been a statistical artefact, especially in consideration of further results. If, as they claimed, 'the 2006 score was partly the result of reforms, policies, strategies, and interventions that were put in place years ago, even a generation ago', then the trend should have continued in 2009 and 2012, but this was not the case. In Mathematics, the 2003–2006 gain was followed by a 20-point loss in 2009 and, once socio-demographic changes are taken into account, a flat trend emerged between 2003 and 2012 (OECD, 2014g). Lenkeit & Caro (2014) demonstrated that the long-term changes in Reading and Mathematics in Indonesia were completely driven by socio-economic and factors that were 'non-malleable' by education policies.

The Mathematics performance of the country was stable also according to TIMSS 1999–2007 (Mullis et al., 2008). TIMSS and PISA are different, but not incomparable. In fact, Wu (2009) showed that once student age and content coverage were taken into account, TIMSS and PISA 2003 shared 93% of the score variance; simple adjustments to sampling and assessment content would be sufficient to cancel the gap between PISA 2003 and 2006 for Indonesia. In summary, the improvements in Reading or Mathematics in Indonesia—and especially the gain between 2003 and 2006—were probably not driven by recent education reforms.

The examples of Indonesia and Qatar suggest that policy alignment may stop at the level of the intentions and not be fully achieved in practice, even when a country has the resources to invest in reforming the system. Moreover, the analyses above support the hypothesis that the impact of country-level education interventions may be small and incremental, rather than large and sudden. Most changes are smaller than a fifth of a PISA standard deviation, although partner countries appear to have experienced larger score fluctuations, which may be due to changes in the socio-economic composition of test-takers.

Multilevel analysis: change between OECD and partner countries over many time points

To model the trends in Figure 4.1 or Figure 4.2, the OECD uses a method referred to in the literature as 'no pooling' (Gelman & Hill, 2007), whereby within-country observations (PISA scores) are regressed separately for each country (OECD, 2012e, 2014g). The method accurately tracks individual trends, but it tends to over-estimate between-country differences and has trouble detecting global trends.

Multilevel or hierarchical modelling (Gelman & Hill, 2007; Goldstein, 2011; Raudenbush & Bryk, 2002), in contrast, is a form of 'partial pooling' analysis that can accurately model both individual and general trends by taking into account between- and within-country similarities. To do this, it assigns a variance component not only to Level-1 units (within-country scores over time), but also to average country achievement (the Level-2 units). For ease of reading, only relevant results are discussed in this section; technical details are available in Appendix 5 and 6.

A model accounting for differences in initial country performance levels and allowing for different slopes in country trends (the random-intercept random-slope Model C in the Appendix 3, Tables 4.C and 4.D) provided a good fit for the data. It confirmed that the between-country variance is much larger than the within-country variance, this latter accounting for about 2–5% of the total variance. It also confirmed that PISA scores change little from year to year: the average effect of time on scores is between +2.4 and +3.3 points per cycle, or +0.6–1.1 point per year depending on the domain. The global effect of time on scores is small because of the opposite trends of OECD and partner countries: while the improvement of OECD economies tends to shrink or even turn negative with time, the improvement of partner countries remains strong and positive.

The between-country variance is decreasing with time (Appendix 3, Table 4.E) and the covariance between intercept and slope is negative. In other words, lower-achieving partner countries improve at a faster rate than higher achieving OECD ones. Altogether, these results corroborate the impression of score convergence, or 'fanning in' (Rasbash, Steele, Browne, & Goldstein, 2012), expressed at the beginning of this chapter. This linear model was also able to capture most of the within-country variance. The fact that 12 years of PISA can be easily summarised with one line suggests that for the most part the effect of policies on scores must be small and progressive, rather than large and sudden.

Model diagnostics (Appendix 6) revealed a curvilinear structure was found in Reading data. This replicates, in a different domain and on a larger scale, OECD findings whereby the rate of change in Mathematics is not constant in some countries (OECD, 2014g, p. 55). The implications of this are discussed in the next section.

The implications of trends linearity, country clustering and changes in between-country variance

What can linearity, score differences between OECD and partner countries and decrease in spread tell us about the use and effectiveness of the EPA?

Linearity in country trends

Linearity, paired with a low within-country variance, implies that either the EPA has not been fully implemented or it has yet to produce any detectable effects: countries simply continue in their set course. This is why the curvilinear trends reported above are interesting: if they continued for another two or three administrations, and if they were not due to score drifting or teaching to the test, they could be indirect evidence of the effectiveness of the education reforms of the early 2000s.

Part of the difficulty in looking for policy effects is that there is little indication in the literature of how long it takes for policies to have an impact, and therefore what time frames to observe. Time frames range from two years up to a generation. Fullan (2000) argued that school districts can improve in 8 years—or perhaps in four (Fullan, 2007, p. 18). Others have claimed that sizeable change can be achieved in an even shorter time span. For instance, Earl, Watson, & Torrance (2002, p. 35) argued that the English National Strategies for Literacy and Numeracy 'made significant changes in primary education throughout England in a remarkably short period of time' (two years; see also Levin et al., 2008, on Ontario). A follow-up report on the English strategy, though, noted that, without pressure and support, the improvements plateaued after four years (Hargreaves & Fink, 2006). Borman et al. (2003) found that sustaining the implementation of comprehensive school reforms produced, after five years, effect sizes that were twice as large as those at the beginning of the intervention. Harris (2011) argued that implementing successful reforms can take ten years, whereas Wedell (2009) stated: 'it is clear that the successful implementation of educational change takes a long time'—perhaps a generation or more (p. 18).

The analyses in these sections suggest that PISA is able to pick up changes in performance only after three or four administrations, on average; therefore, effective reforms introduced between PISA 2000 and 2003 are likely to produce observable effects only in PISA 2009 or 2012. Of course, large variations in country scores can be seen even between two administrations, which could be evidence of countries implementing 'quick fixes' that are immediately absorbed by the system because of a lack of political and financial support (Harris, 2011; Hopkins, Stringfield, Harris, Stoll, & Mackay, 2014). However, the .99 correlation between Mathematics 2003 and 2006, the case of Indonesia and evidence from TIMSS suggest that single large fluctuations are more likely to be statistical artefacts.

The different magnitude of score changes among OECD and partner countries

It is more difficult to interpret why score changes have been larger among partner countries. The compelling narrative promoted in international reports is that, perhaps, these education systems were more malleable to the reforms auspicated by the OECD or they were more ready to implement them, unlike 'complacent' countries living off traditions and past reputation (Schleicher, 2013f). Turkey was one of the lowest achievers but managed to improve its performance—and that of its students in the lowest percentiles—with the contribution of a curricular reform reflecting 'the assumption, on which PISA is based, that schools should equip students with the skills needed to ensure success at school and in life' (OECD, 2014g, p. 123).

This narrative, however, does not explain the score drop in Malaysia, a country that has directed considerable attention to pre-primary education, teacher training and quality monitoring (all featuring in the EPA) well before its first participation in PISA (UNESCO-IBE, 2011b). In a later section, the reliability of PISA over time and its robustness to confounding factors are investigated, as are alternative explanations of the difference in the size of score changes between OECD and non-OECD countries.

The decrease in between-country variance

The reduction in spread might be seen as evidence of globalising forces in education: countries align their policies to a common standard—which in this case may be the OECD standard summarised by the EPA—and this translates into more homogeneous outcomes.

There is little empirical research on this phenomenon using data from international LSAs. Rutkowski & Rutkowski (2009) tried to detect curricular convergence in mathematics by looking at the item response patterns of 16 countries in TIMSS 1995–2003. The authors found that 'rather than students' responses becoming more similar over time, regional clusters remained intact over the three administrations' (p. 148), suggesting little convergence on a global scale. Similar findings were reported by Dudaite & Elijio (2008) looking at TIMSS Mathematics 1995 and 2003; they also added that between-country correlations increased in those eight years, so that countries were 'drawing closer to each other in their mathematics profile' (p. 6), but the size of this effect was very small. In fact, an increase in Pearson's *r* does not imply that countries were drawing closer; rather that their trends were becoming more stable.

There are two issues with attributing the reduction of country variance to country alignment with the EPA or to the PISA curriculum. Firstly, the standard error (SE) on the measurement has decreased in partner countries by twice the amount it did in OECD countries, and currently the values for the two groups are roughly the same (own calculation using data

from PISA 2012, but see also OECD, 2014h, Table C.1). In other words, PISA is becoming more accurate, especially in relation to the performance of non-OECD countries, which may explain the between-country variance reduction. Secondly, this view does not clarify why the results of the OECD cluster are worsening with time, as it emerged from the blue lines in Figure 4.1 and Figure 4.2 and was confirmed by the multilevel models. Why would any alignment with the EPA have a positive effect in partner countries and a negative effect in OECD countries?

Trends in equity

The OECD defines equity in terms of fairness and inclusion: 'Equity as *inclusion* means ensuring that all students reach at least a basic minimum level of skills. […] Equity as *fairness* implies that personal or socio-economic circumstances, such as gender, ethnic origin or family background are not obstacles to educational success' (OECD, 2012c, p. 15, emphasis in the text).

Inclusion is often measured negatively in terms of the percentage of students achieving below the lowest PISA proficiency level (1 or 2 depending on the cycle); it can be argued, however, that raising the scores of the lowest achievers might make a system "technically" equitable according to PISA without narrowing social gaps. For example, the success of Poland in PISA was attributed to the 1998–1999 reform delaying entry to vocational education by one year: students no longer had the option to enter vocational education after eight years of primary education, but only after six years in primary followed by three years in lower general secondary education (OECD, 2011h; UNICEF, 2012). As a result, between-school variance and the percentage of low achievers decreased. Yet, the relationship between students' socio-economic background and scores did not: more advantaged student had the same edge on more disadvantaged students before and after the reform. In other words, gains in PISA were not paired by social gains: 'the negative effect of the tracking system was simply postponed by one year' (OECD, 2011h, p. 25).

A stronger measure of equity, in this author's opinion, makes use of the concept of "fairness", which generally refers to equality of opportunity regardless of socio-economic and demographic circumstances (Roemer, 1998). One possible measure of this is the socio-economic gradient. The "socio-economic gradient" is the regression line (or curve) of student scores on their composite index of economic, social and cultural status (ESCS). It has different attributes (see OECD, 2013e, p. 35; Willms, 2006), but the most important are its Strength and its Slope (capital letters are used to differentiate them from general statistical terms).

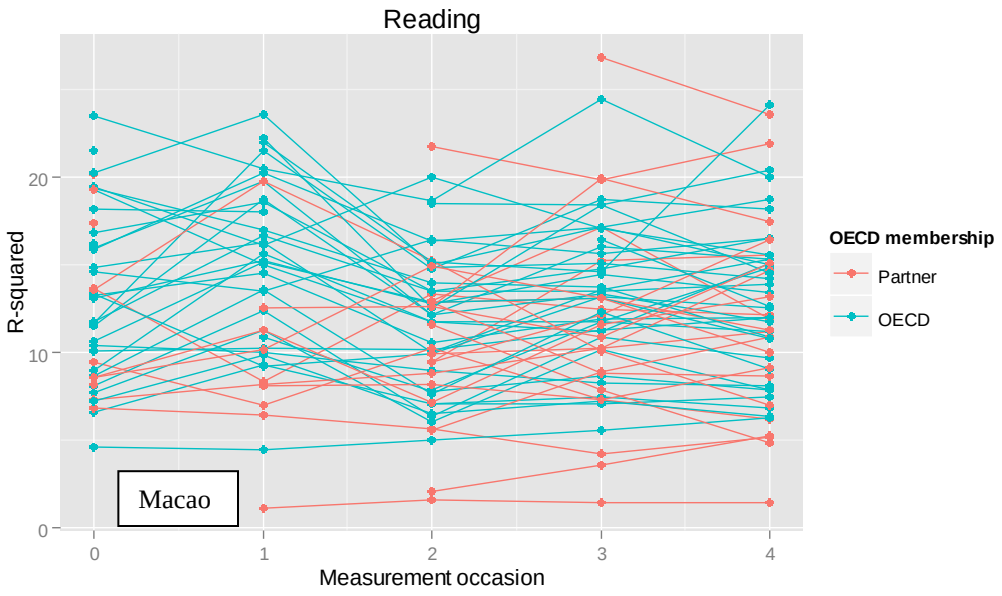
The "Strength" of the socio-economic gradient is the amount of score variance captured by the regression line (in statistical terms, the R^2): the higher the "explained" variance, the

stronger the ESCS-score relationship, and the less equitable a system. The "Slope", instead, refers to the extent to which differences in socio-economic background are associated with differences in outcomes, with steeper slopes being related to greater disparities. Theoretically, using either one of the two measures in statistical analyses should yield the same results, because if ESCS and PISA scores were converted to standardised scores (Z-scores), the correlation between them would be the Slope of the line and the R^2 the Strength. Nevertheless, it was decided to follow the OECD procedure and run analyses using both attributes as the outcome variable.

Figure 4.5 and Figure 4.6 map, as an example, trends in Strength and Slope using Reading as achievement measure. Unlike trends in performance, trends in equity are less linear. Fewer differences between OECD and partner countries can be detected and when they are more visible (e.g., in the Slope), this can be explained by the higher average scores of OECD members rather than substantial differences in equity. In fact, the rates of change look similar. The between-country variance seems to increase rather than decrease.

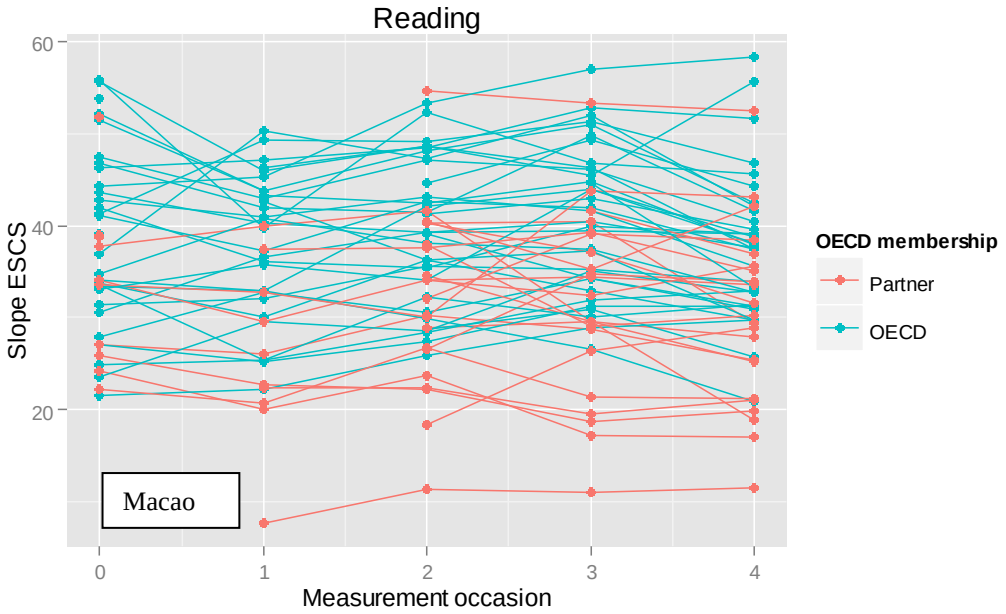
Notice that there is an inverse relationship between equity and the values on the *y*-axes: the *higher* the socio-economic gradient, the *lower* the equity (the higher the inequality). According to PISA, Macao is the most equitable country, at least in terms of educational opportunities. On an international level, the average Strength accounts for between 11% and 16% of the country score variance depending on the domain and the year, whereas the Slope for 34–38 points.

Figure 4.5: trends in the Strength of the socio-economic gradient for Reading.



Source: own elaboration using a new dataset of equity measures derived from PISA data. The unit on the *y*-axis is "percentage of the 'explained' variance".

Figure 4.6: trends in the Slope of the socio-economic gradient for Reading.



Source: own elaboration using an original dataset of equity measures derived from PISA data. The unit on the *y*-axis is "PISA points".

To detect general trends, separate single-level and multilevel models were fitted, following a similar approach to that used in the previous section to estimate trends in performance.

Table 4.F in Appendix 3 reports the initial level and average change of the Strength and Slope of the socio-economic gradient in different countries. Thirty-eight percent of PISA countries experienced statistically significant changes in equity until 2012. In most cases only the Strength or the Slope reached the significance level, probably because of the large SE. In countries where changes in both Strength and Slope were statistically significant the direction of the trend was the same and opposite trends were never observed, which corroborates the validity of the methodology (remember that the Strength and the Slope are supposed to be two sides of the same coin).

Table 4.3: Trends in the Strength or the Slope of the socio-economic gradients.

Strength →	*Increase*	*No change*	*Decrease*
↓Slope			
Increase	Romania, Russian Federation, Taiwan	Finland, France, Iceland, Korea, New Zealand, Portugal, Qatar, Slovak Republic, Uruguay	
No change	Israel, Latvia	ALL OTHER COUNTRIES	UK, Netherlands, Thailand
Decrease		Argentina, Slovenia, USA	Germany, Jordan, Liechtenstein, Mexico, Turkey

Source: own elaboration using an original dataset of equity measures derived from PISA data

Table 4.3 shows that, on a global level, the situation is rather balanced: in most countries equity levels are stable; 14 countries have experienced an increase in inequality (trends in either the strength or the slope of the socio-economic gradients are positive), and 11 countries have experienced a decrease. Stronger evidence of higher inequality is available for Romania, the Russian Federation and Taiwan, whereas the opposite holds for Germany, Jordan, Liechtenstein, Mexico and Turkey. Liechtenstein in particular has lost on average 9–12 percentage points in Strength since the first PISA administration.

Multilevel models were also fitted. Results are shown in Appendix 3, Table 4.G and Table 4.H. The balanced situation found through single-level analysis was mostly confirmed: inequality levels appear to be stable. There is some evidence that the Slope might be decreasing in Science, but the magnitude of the change is small: while more advantaged students used to score 31.5 Science points more than the average student in 2006, by 2012 the gap had been reduced to 27 points. Likewise, there is some evidence that the Strength of the socio-economic gradient in OECD countries in Reading, which was on average 14% of the total score variance in 2000, decreased to about 11% in 2012. Unlike trends in performance, however, the between-country variance is increasing over time (Appendix 3, Table 4.I), which means that countries are becoming more different in their inequality levels.

Overall, the stability of equity levels in both OECD and partner countries suggests that, if countries are implementing policy measures following the EPA, these are having little effect on a global scale. For each country experiencing a decrease in inequality there is another experiencing an increase, and another two showing no change. It is more difficult to interpret the widening spread between countries; given the varying number of countries participating to each administration and the small magnitude of the change, this may simply reflect random fluctuations.

Second step: the limits of education reforms

The evidence gathered thus far suggests that trends in achievement and equity are highly stable, and therefore that any effect of education policies is either very small or has yet to be seen. Year-on-year fluctuations could be due to random errors, whereas convergence in PISA performance is paired with divergence in equity measures. Moreover, it is possible that factors other than education policies underlie the observed trends. It was shown earlier that, in the case of Qatar and Indonesia, changes in the socio-economic and demographic composition of the student population explained most changes in PISA outcomes. Accounting for these variations reduces the increase of Qatar by 20 points, Chile by 19, Brazil by 11, Turkey by 11, and it makes some of these improvements statistically non-significant (own calculation from the PISA database, see OECD, 2014g, p. 188, for the impact on annualised change).

These findings raise a question about the extent to which country performance may be affected by factors unrelated to education policies. Answering this question can set an empirical limit on the ability of education policies to influence PISA outcomes. If factors non-malleable by education policies were consistently found to account for, say, 50% of country results, then education policies could only be said to make "half" the difference.

In this section, the limits of education reforms are explored by contrasting the effects of three policy outputs (economic productivity, equitability and access to pre-primary education) on PISA outcomes. All three factors have a particular relationship with PISA scores and a different place in the OECD narrative.

It is known that country economic productivity (generally expressed in terms of GDP or GDP per capita) correlates with PISA scores, though the OECD tends to downplay the magnitude of the correlation. It is known that more socially equitable countries have better social outcomes, and the OECD argues that more educationally equitable countries have better educational outcomes. Finally, it is known that the quality and quantity of early childhood education has a positive impact on educational trajectories and later life outcomes, and the OECD has recently presented PISA evidence that access to pre-primary education correlates with higher scores in Reading, Mathematics and Science.

What is not known is whether *changes* in the three factors over time are associated with changes in PISA outcomes, and if they are, whether their relationships with PISA are of comparable magnitude. In this research, it is posited that any detectable "effect" on PISA scores should be larger for the factor that is more malleable by specific education policies (access to pre-primary education), smaller for the "generic" educational factor (equitability) and even smaller for the non-educational one (productivity).

Because of the high aggregation level and the few time points, even the largest correlation is expected to be of moderate magnitude. Recent studies of educational effectiveness suggest that schools are more 'decoupled' from education policies than previously thought (Scheerens et al., 2015), thus 'the impact of variables that are prone to manipulation by policy makers is quite modest' (p. 33). Under this theoretical framework, 'sustained educational policies can bring about considerable societal changes' but their short-term impact is limited (Scheerens et al., 2015, p. 33; also Snyder, 2013).

The impact of economic productivity on PISA trends

One of the most commonly-used measures of country productivity is GDP. The gross domestic product has an ambivalent role within the OECD narrative. The OECD reports that 'countries with higher national income[13] tend to perform better' (OECD & UNESCO-UIS, 2003a, p. 111;

[13] Represented by GDP in this context, even though "GDP" and "income" are not synonymous.

OECD, 2001a, p. 92, 2004b, p. 100, 2007b, p. 59, 2010g, p. 34), but it also admonishes its audience that 'the countries that are the strongest performers in PISA are not the wealthiest' (OECD, 2012f, p. 1).

The latest statement belongs to a narrative of possibility whereby high achievement is within reach of all countries in spite of their contextual conditions, which is necessary to make PISA appealing to developing countries. The problem is that the narrative of possibility conflicts with another position the OECD holds: the belief in human capital theory. Human capital theory states that education and wealth are related, but this relationship corroborates the narrative only insomuch as human capital (measured by PISA outcomes) 'causes' economic growth (Hanushek & Wößmann, 2012).

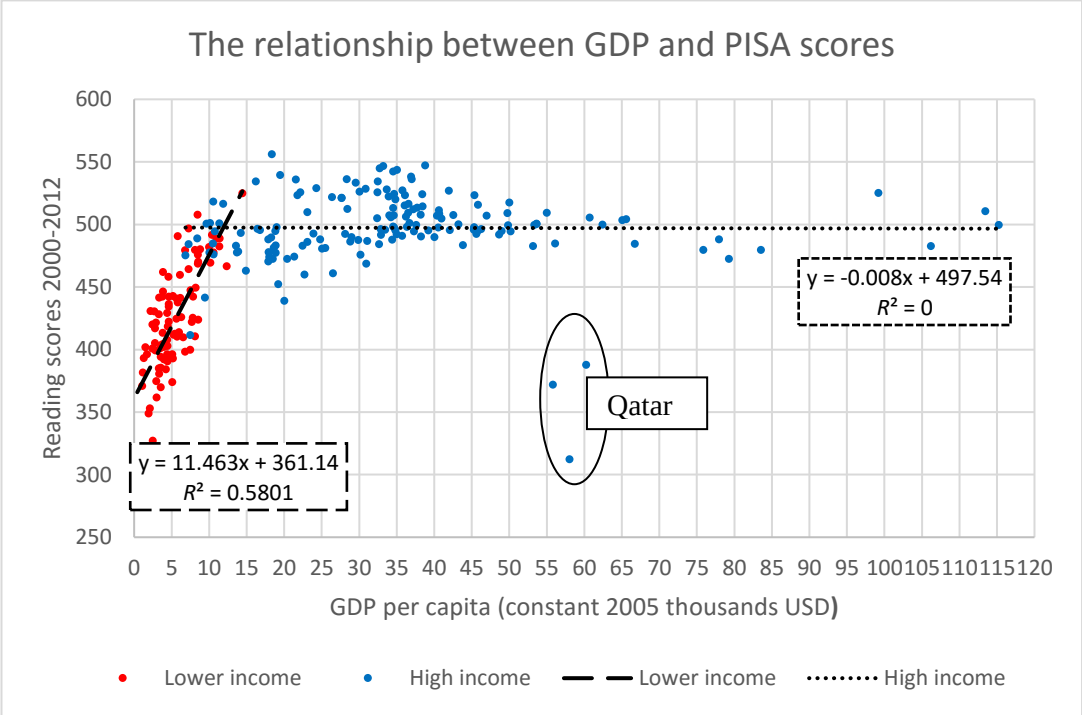
If the reverse were true, in fact, one could think that country performance in PISA had more to do with economic rather than educational policies, which is an issue when the target audience of the survey are education stakeholders. Could Argentina have scored 44 points less in 2006 than it did in 2000 as a result of its economic crisis of 1998–2002? The OECD did not investigate this question.

One strategy the OECD adopts to hold the narrative and the theory together is to acknowledge but downplay the correlation between GDP and PISA scores. According to OECD figures, the extent to which GDP per capita "explains" country differences in PISA scores has dropped from 43% in 2002 to 21% in 2012 reaching even lower values in-between (OECD & UNESCO-UIS, 2003a; OECD, 2014g). If one takes into account how the data was calculated or presented, however, a different story emerges.

In 2002, simply 'excluding Luxembourg, which is an extreme outlier due to its high per capita income, [increases] the overall correlation coefficient across all participating countries […] to 60 per cent' (OECD & UNESCO-UIS, 2003a, p. 111). Following the same approach and removing the outliers from the 2012 scatterplot increases the variance "explained" (the R^2) from 21% to 59%. As another example, in PISA 2009, the OECD-reported figure of 6% (OECD, 2010g, p. 14) was computed using data from OECD countries only, which tend to cluster in the flatter part of the graph.

In general, it is true that GDP levels and PISA outcomes are only partially related: using all available data and including outsiders, the R^2 has been stable at 22%–32% since PISA 2000. It is also true that this figure masks important differences between richer and poorer countries (Figure 4.7).

Figure 4.7: The relationship between GDP per capita (thousands USD), income and PISA scores (Reading) 2000–2012



Source: own elaboration using PISA data

Figure 4.7 displays the relationship between GDP per capita (expressed in constant 2005 US Dollars) and Reading scores in lower and high income countries[14] between 2000 and 2012. Outcome data have been pooled; therefore, countries are represented by more than one dot, as the example of Qatar shows in the picture.

There is a strong correlation between wealth and PISA outcomes in lower-income countries, with GDP accounting for 58% of the score variance on average across administrations. Past this threshold, the line becomes flat and GDP is no longer associated with Reading scores in high-income countries. This difference has been evident since the first PISA administration, as Table 4.4 shows.

Table 4.4: The relationship between income and PISA scores (reading) over time, by country income.

	2000	**2003**	**2006**	**2009**	**2012**
Lower income	0.76 (16)	0.73 (13)	0.83 (24)	0.76 (23)	0.64 (18)
High income	-0.01 (23)	0.07 (26)	-0.09 (32)	-0.07 (38)	0.12 (41)

Source: own elaboration using PISA data. The first value in the cell is Pearson's *r*. The number between brackets is the sample size.

[14] The World Bank divides economies into four groups according to their gross national income (GNI) per capita: low, lower-middle, upper-middle, and high income. In this research, "lower income" includes countries from the first three groups (low, lower-middle and upper-middle). The "high income" category corresponds to a GNI per capita above $9,265 in 2000 up to above $12,615 in 2012 (The World Bank, 2014).

Interestingly, the strength of the correlation appears to have weakened in 2012 for lower-income countries and strengthened for high income ones, although the small sample sizes prevent firm conclusions.

The fact that PISA correlates with GDP more strongly in poorer countries is consistent with the literature (Binder, 2009; Brist & Caplan, 1999). However, this does not explain PISA trends, only the relative position of countries along the performance distribution. The question is whether changes in GDP may be associated with changes in scores and—given that many strong improvers are low-performing, lower-income partner countries—whether the relationship plays differently between OECD and non-OECD countries.

Recent research suggests that this is a real possibility. Permani (2009) found evidence that GDP growth can foster educational outcomes in a range of East Asian countries, regardless of their income: low (Cambodia), lower-middle (Thailand), upper-middle (Malaysia), or high (Singapore). Mazumdar (2000) found that adult literacy rate was a precondition for higher GDP per capita in high-income countries but a consequence of it in lower-income ones. Cracolici, Cuffaro, and Nijkamp (2010), though, reported for a sample of 64 countries observed between 1980 and 1999 that 'there exists only a unidirectional relationship from *gdp* to *li* [literacy rate], i.e. *gdp* precedes *li*' (p. 349).

To model the relationship between changes in GDP per capita and changes in educational outcomes, the multilevel growth model used earlier for trends in performance and equity was expanded to include country average and change in GDP per capita as regressors (see Appendix 5 for details). Clearly, this relationship is mediated by the introduction of education policies that were made possible by higher productivity, and this takes an unspecified amount of time. Therefore, all changes in GDP taking place during the lifetime of each PISA cohort were considered, which corresponds to regressing PISA outcomes on economic trends from 1985 until 2012.

Table 4.5 reports, as an example, the results from the period 1985–1997. A regression over this period compares the different economic conditions in which each of the five successive Reading cohorts was born, and models whether changes in these conditions were related to better or worse student scores 15 years later. Trends for other periods were also modelled (i.e., GDP when PISA students were one year of age, two, three and so on), but the outcomes were similar for all time frames and all domains.

Table 4.5: A multilevel model of Reading scores on GDP per capita 1985–1997, when the five successive PISA cohorts were born, for all countries and for a subsample of lower-income countries.

Reading	**All countries**			**Lower-income countries**		
	Estimate	*Error*	*DF*	*Estimate*	*Error*	*DF*
(Intercept)	407.89***	(10.01)	162	367.69***	(14.60)	53
Δ GDP	0.08	(0.41)	162	-0.07	(2.60)	53
Time	4.90***	(1.24)	162	5.98***	(1.24)	53

OECD	66.87***	(12.23)	48			
AVG GDP	0.69**	(0.32)	48	11.88***	(2.88)	18
*Time*OECD*	-4.13***	(1.50)	162			
σ2u0	1466.04	(994–2276)		981.73	(455–2116)	
σ2u1	10.40	(3.94–27.92)		7.87	(0.65–95.35)	
σu01	-75.031			-43.037		
σ2e	98.57	(76.34–127.27)		115.92	(73.33–183.25)	

***p <0.01, **p <0.05, *p <0.1. *Source*: own elaboration using PISA data. σ^2_{u0} is the variance of country intercepts (the variability of PISA scores upon first administration); σ^2_{u1} is the variance of country slopes (the variability of the rate of change in PISA scores); σ^2_{u01} is the covariance of intercepts and slopes; σ^2_e is the within-country variance (the variability PISA scores around individual country trends). "Δ GDP" captures the effect of a within-country change in $1,000 per capita GDP on PISA scores over and above the effect of *Time*, and regardless of a country's mean per capita GDP and OECD membership (note the interaction between *Time* and *OECD*, capturing the fact that OECD countries as a group have improved less in PISA over time than partner countries). "AVG GDP" measures the extent to which a $1,000 between-country difference in GDP per capita is associated with PISA outcomes. The subsample of lower-income countries (n = 29) on the right part of the table is composed of countries whose GDP per capita was below $17,000 in 2012. This is higher than the World Bank definition of "lower income", but close to the point at which the scatterplot in Figure 4.7 "bends". This amendment was necessary to increase the sample size. Given that most countries in the sample were not OECD members, the OECD variable was excluded from the model.

On a global level, changes in GDP are never associated with changes in Reading, Mathematics or Science, regardless of whether they occurred very close to the time when the students sat the test or 15 years earlier. Instead, average GDP is consistently and significantly related to PISA scores, confirming the findings from the correlational analysis at the beginning of this section. When all countries are taken into consideration, the magnitude of this relationship is small: a between-country difference of $10,000 is associated with a difference in Reading of 5–7 points. It was argued that the GDP–PISA correlation varies by country income, however, and in fact the analysis carried out on a subset of lower-income countries suggests that in these contexts wealth "matters" more: the same $10,000 difference in GDP per capita corresponds to 100–120 points in Reading. This is 20 times the relationship existing in high-income countries.

Other analyses (not shown) were carried out to see whether the correlations would be higher for more policy-malleable outputs such as the expenditures per student in primary or secondary education, in lower- and high-income countries. For comparability reasons, the analyses only considered the quantity of money spent for education but not how it was allocated.

All findings pointed to the same conclusions: in richer and poorer economies alike, changes in GDP or expenditures are not followed by related changes in PISA outcomes. Yet, on average, GDP per capita and spending per student strongly correlate with achievement levels, at least until a threshold which could be placed between $7,000 and $17,000 for GDP, and between $2,000 and $5,000 for student expenditure. There was also some evidence that, among lower-income countries and in recent years, greater changes in GDP corresponded to smaller changes in PISA. This may have captured the threshold effect on some fast-developing

countries moving left to right in Figure 4.7 and therefore entering the flat part of the graph. Some exceptions are Chile or Poland, which have moved to the 'high income' category while increasing their PISA scores by more than 30 points.

That a strong correlation could be observed between average PISA outcomes and average GDP, but not between changes in the two variables, might seem paradoxical, but this is probably a limitation of the model employed. GDP tends to grow over time in all countries, whereas PISA scores are more variable. By matching a general tendency to rapid fluctuations, the model cannot pick up those small departures from the generalised GDP growth that might instead make a difference to PISA scores and which returns zero correlations. ARIMA models could be better suited for future analyses.

The relationship between educational equity and PISA trends

The OECD position is clear: 'countries do not have to sacrifice high performance to achieve equity in education opportunities' (OECD, 2013e, p. 3) and this is because, 'if some countries are able to achieve high levels of learning outcomes for students from different backgrounds, the implication for others is that it is feasible to meet both equity and quality goals' (OECD & UNESCO-UIS, 2003a, p. 164). Secondary analyses of PISA data support this position (Condron, 2011; Willms, 2006), and in general there is little doubt in the academic world that equity and achievement are compatible.

At times, however, the OECD has taken a stronger stance on the matter: 'the evidence is conclusive: equity in education pays off.' (OECD, 2012c, p. 14). Therefore, '*investing* in […] education for all, and in particular for children from disadvantaged backgrounds, is both fair and economically efficient' (p. 9, emphasis added). This is a substantially different statement: it is one thing to say that equity does not hurt achievement but another to claim that high achievement "needs" an equitable system.

The literature supporting this "stronger" statement generally covers the intergenerational consequences of inequality, such as differential educational attainments between parents and offspring (S. E. Black & Devereux, 2010; Bol & van de Werfhorst, 2013; Pfeffer, 2008; Salverda, 2011). Educational literature agrees that increasing inequality does not improve performance (e.g., Duru-Bellat & Suchaut, 2005), whereas influential books such as *The Spirit Level* (Wilkinson & Pickett, 2009) have shown that inequality has a detrimental effect on societies over long periods of time: countries that became less equitable had increased crime rates and those which became more equitable had decreased crime rates. Could a similar relationship be observed, and in a short span of time, in terms of countries becoming more or less educationally equitable and experiencing increases or decreases in PISA outcomes?

Earlier in this chapter, it was shown that country inequality levels (as defined by PISA in terms of Strength and Slope of the socio-economic gradient) have remained stable, which casts some doubt on the ability of equity to have played a role in the observed performance trends. Moreover, Teltemann (2014) studied whether the educational reforms implemented after PISA were effective in promoting both high achievement and more equity. She found that the implementation of more or less equitable policies was generally uncorrelated with PISA outcomes, with the exception of non-targeted ability grouping, whose increase at the country level was also associated with an increase in inequality.

These findings are followed up in the current section, which explores the relationship between country changes in equity and PISA outcomes with correlational and multilevel analysis.

In the correlational approach, countries which experienced a significant change in the Strength or Slope between the first and the last available administration were selected. It was found that in over half the cases, significant changes in the socio-economic gradient corresponded to non-significant changes in PISA scores: observable changes in equity had not had any detectable effect on outcomes. These cases were excluded from further analyses.

The remaining two sets of data, comprising countries with statistically significant changes in both socio-economic gradient and PISA scores, were correlated (the decision to restrict the dataset was so that only "clear" cases with changes in both variables were analysed). The results are presented in Table 4.6. Results suggest that changes in both the Strength and the Slope correlate moderately and negatively with PISA scores, meaning that increases in inequality may indeed be associated with decreases in scores.

Table 4.6: Correlation between country change in Strength or Slope between first and last comparable PISA participation and changes in their PISA scores

Reading		**Mathematics**		**Science**	
Strength	*Slope*	*Strength*	*Slope*	*Strength*	*Slope*
-0.26 (8)	-0.25* (30)	-0.11 (8)	-0.42** (27)	-0.16 (5)	-0.22 (24)

***p* <0.05, **p* <0.1 (one-tailed). *Source*: own elaboration using PISA data. The first value in the cell is Pearson's *r*, the number between brackets is the sample size.

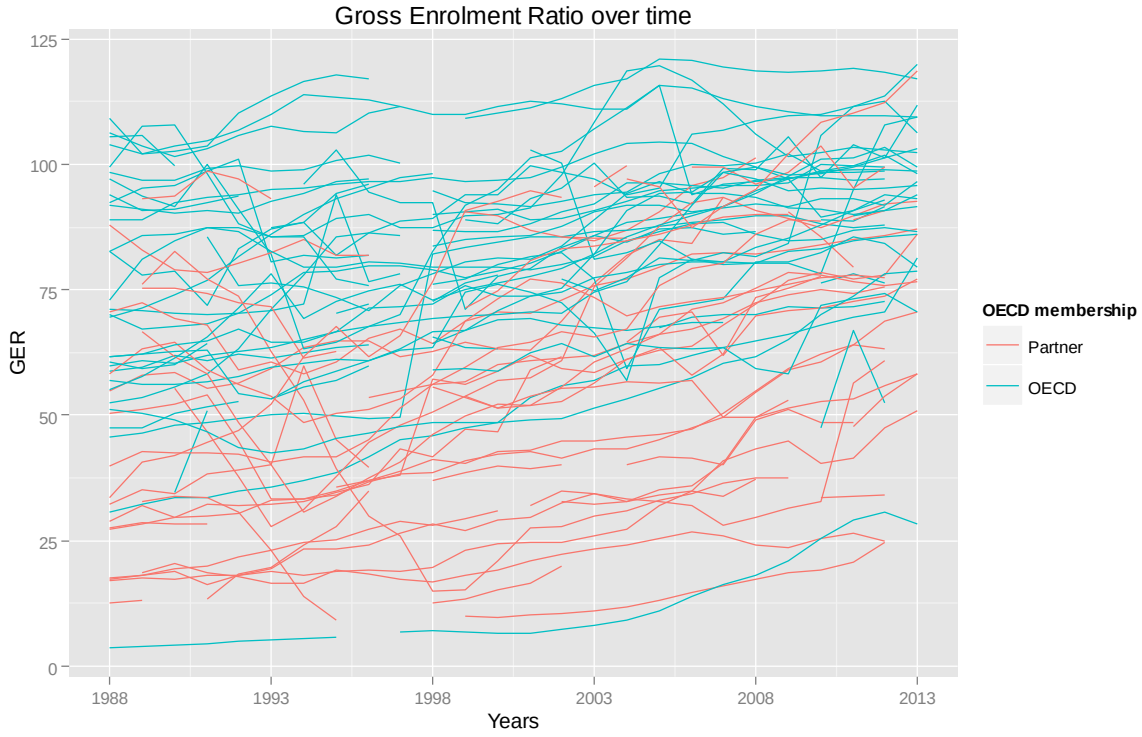
Multilevel models were also employed, regressing PISA outcomes on the Strength and the Slope of countries' socio-economic gradient and using the same approach employed in the rest of the chapter. Contradicting the correlational analysis, however, no evidence was found (Appendix 3, Table 4.J) that average levels or trends in the socio-economic gradient were associated with changes in achievement. Taken together, these contrasting findings provide more support to the OECD's "weaker" argument (more equity does not translate into worse achievement) than its "stronger" one (more equity translates into better achievement). This

conclusion applies particularly to the Strength of the socio-economic gradient, which is 'PISA's main measure of equity in education outcomes' (OECD, 2013e, p. 35).

Increased access to pre-primary education and changes in PISA scores

Figure 4.8 maps trends in gross enrolment ratio (GER) in OECD and partner countries between 1988 (when the first PISA cohort was aged 4) and 2013. Despite some large drops between 1988 and 1995–1998, and many within-countries fluctuations, the global trends have been on the rise.

Figure 4.8: Gross enrolment ratio in OECD and partner countries 1988–2013



Source: own elaboration using UNESCO data

The average GER was 59.5% in 1988 (SD = 29, Median = 59.9%, median absolute deviation [MAD] = 26.2, n = 49, OECD = 31) and 86.9% in 2013 (SD = 21.79, M = 92%, MAD = 12.9, n = 36, OECD = 24). OECD and partner countries have increased enrolment ratios by the same amount, on average: slightly more than one percentage point per year (Appendix 3, Table 4.K). An analysis of variance showed that a curvilinear model provides a better fit for the data, meaning that the rate of change has been increasing. However, the linear and curvilinear models in Appendix, Table 4.K differ by only about one percentage point when it comes to predicting the average GER across OECD countries in 2013.

At the time of the first PISA administration, of the 53 countries with a valid GER value, only 14 had a coverage below 50% and they had all increased it by PISA 2012 (data not

shown). By 2012–2013, only two countries had a GER lower than 50% (Turkey and Tunisia). Although the gross enrolment ratio is probably an overestimation of actual enrolment, there does not seem to be a crisis of access to pre-primary education. This suggests that the OECD recommendation (increase overall time spent in education, through policies expanding access to pre-primary education, targeting enrolment or changing the age of entry into school) is grounded only insomuch as it refers to bridging the participation gap between advantaged and disadvantaged children, rather than to widening overall participation.

Increased access to pre-primary schooling is a very specific education policy whose effectiveness is strongly backed by the literature. Therefore, it was a better candidate than GDP and country equity levels for uncovering significant relationships with PISA outcomes. To evaluate the impact of increase in GER on PISA outcomes, a multilevel model was fitted using GER data from the 1988–2000 period, when each successive PISA Reading cohort was aged 4. Results are presented in Table 4.7.

Table 4.7: A multilevel model of the relationship between mean and change in pre-primary GER and PISA

	Reading			**Mathematics**			**Science**		
	Estimate	*Error*	*DF*	*Estimate*	*Error*	*DF*	*Estimate*	*Error*	*DF*
(Intercept)	377.06***	(12.84)	171	370.93***	(13.80)	135	403.18***	(15.26)	90
Δ GER	0.35***	(0.13)	171	-0.01	(0.14)	135	0.00	(0.15)	90
Time	4.72***	(1.43)	171	6.14***	(1.77)	135	4.81**	(1.93)	90
OECD	70.19***	(13.03)	57	61.90***	(14.10)	57	43.72***	(12.84)	48
AVG GER	0.60***	(0.21)	57	0.80***	(0.22)	57	0.65***	(0.24)	48
Time × OECD	-4.79***	(1.70)	171	-6.51***	(2.09)	135	-3.53	(2.17)	90
σ^2_{u0}	1695.88	(1116–2576)		2083.66	(1418–3061)		1366.24	(909–2052)	
σ^2_{u1}	21.40	(9.76–46.90)		41.48	(21.80–78.89)		37.08	(20.33–67.62)	
σ_{u01}	-127.919			-205.433			-70.480		
σ^2_e	92.07	(70.67–119.90)		44.52	(31.85–62.22)		23.42	(15.18–36.13)	

***p <0.01, **p <0.05, *p <0.1. *Source*: own elaboration using UNESCO and PISA data. As in Table 4.5, σ^2_{u0} is the variance of country intercepts; σ^2_{u1} is the variance of country slopes; σ^2_{u01} is the covariance of intercepts and slopes; σ^2_e is the within-country variance.

Average country GER is significantly related to PISA outcomes, though the size of its effect is much smaller compared to that of OECD membership: a ten percentage-point difference in pre-primary enrolment is associated with a 6–8-point gap in PISA (but see below). Also noteworthy is that increased enrolment (captured by ΔGER) is linked with changes in Reading twelve years later. Specifically, a ten percentage-point increase in GER corresponded to an increase of 3.5 Reading points, over and above any impact of other policies captured by the time variable, but no link was found with Mathematics or Science.

Even if these coefficients seem small (increasing coverage from 50% to 100% would theoretically increase PISA scores by about half a standard deviation), it should be borne in

mind that these relationships were detected after a decade. And, to put things in perspective, it is a similar correlation to that associated with a $10,000 difference in GDP per capita (or $1,000 in developing countries).

The analysis was repeated separately on lower- and high-income countries (see footnote 14 for a definition) to see if average or changes in GER 'mattered' more to poorer economies. The country sample could be split into two similar-sized categories (n = 29 for lower-income countries, 10 of which were OECD members; n = 28 for high-income countries, 24 of which were OECD members). As expected, OECD members were underrepresented in one group and overrepresented in the other. The results are presented in Appendix 3, Tables 4.L and 4.M.

The average GER coefficient is positive for lower-income countries and not significantly different from zero for high income countries. This means that the effect captured by the full sample model in Table 4.7 was completely driven by developing countries, in which there is almost a 1-to-1 relationship between average PISA scores and average GER: a 10-point difference in GER corresponds to 9–12 PISA points. Interestingly, however, the effect of change in GER on Reading scores is of similar magnitude between the two groups. This suggests that policies affecting access to pre-primary education could be equally valuable to developing and developed economies.

Altogether, these findings provide empirical evidence on an international scale that increasing access to pre-primary education may indeed be beneficial for country performance in PISA, as argued by the OECD. Moreover, they suggest that if any policy effects can be detected through PISA, they might be able to produce up to a 0.3–0.4-point increase in PISA outcomes per percentage point increase in the indicator, though most policies would probably have a substantively smaller impact.

Third step: country alignment with the EPA and outcomes

Results from the previous sections provide a mixed picture about the equity and performance of PISA participants. The only similarity between the two components is that their trends are highly stable. That left aside, OECD countries might be becoming more equitable but their overall performance is decreasing; the between-country variance appears to decrease when measured in terms of achievement and increase when measured in terms of equity.

Most of the time, there is no relationship between changes in inequality and in performance, but when changes are large enough to be detected, they are negatively correlated: countries that reduced their inequality level the most have also experienced the largest score increases. Wealthier countries tend to do better at PISA, yet faster developers are not the strongest improvers. It was also possible to detect the indirect effect of a very strong policy,

the widening of access to early childhood education, on PISA outcomes in lower- and high-income countries.

Thus far, evidence of policy alignment and effectiveness has been general and largely inferential. To reach a greater level of specificity, additional data were collected from experts in different countries on recent national education reforms.

Expert questionnaire

A questionnaire (Appendix 3, Table 2.C) was administered via email to 91 academic and non-academic experts, sampled according to their professional position and presumed knowledge of both PISA and national education policies. The question was: 'To what extent have your government's education policies been in line with the OECD's advice, over the last decade?' Forty experts responded, and their answers (A = Not at all; B = Very slightly; C = To a modest extent; D = Quite a lot; E = Completely) are summarised in Table 4.8 (one respondent per row).

Table 4.8: expert rating of country alignment with the EPA over the last decade

Country	Rating	Country	Rating	Country	Rating	Country	Rating
Australia	C	Denmark	D	Norway	C	Taiwan	D
Belgium (French)	C	Finland	C	Peru	C	Taiwan	C
Brazil	B	Germany	C	Portugal	D	Thailand	B
Brazil	C	Germany	C	Qatar	D	Thailand	C
Bulgaria	D	Hong Kong	D	Russian federation	C	UK	D
Canada	D	Indonesia	D	Russian federation	D/C*	UK	D
Canada (Ontario)	D	Ireland	D	Russian federation	C	UK	A
Chile	C	Japan	C	Slovenia	C	UK	D
Czech Republic	D/C*	Jordan	C	Spain	C	UK (Eng.)	B
Denmark	D	Latvia	C	Sweden	B	US	B

*The answer was conditional to the policy being written or implemented. This is part of the accuracy that had to be sacrificed to keep the stem as simple as possible. *Source*: own elaboration using data from the expert questionnaire.

There is substantial agreement among experts, with 85% indicating that governmental policies in their country were 'to a modest extent' or 'quite a lot' in line with the EPA. The fact that most respondents chose the central option and almost no one went for either of the extremes is unsurprising, given the general nature of the question and the fact that in each country, some policies had probably been informed by the EPA and some had not.

The distribution of responses in options B and D, instead, is very interesting. Only five experts thought that policies in their country were 'very slightly' in line with the OECD policy advice, against 14 who thought they were 'quite a lot' in line. This asymmetry shows that the experts in the sample were more ready to commit to a stronger than a weaker influence of the

OECD in national education policies, suggesting that EPA-aligned reforms might be more common than non-aligned ones.

Another noteworthy finding is that expert ratings from the same country generally agreed. An exception was the response from the UK: three experts stated that policy alignment was strong, one that it was non-existent and one that it was very weak. Unfortunately, with only one question standard techniques to assess inter-rater agreement such as Krippendorff's α (Dorussen et al., 2005) could not be used, and the fact that the UK comprises different countries with independent education systems did not help. The variability in these responses illustrates the difficulty of summarising complex information (such as ten years of policy alignment) reliably. This is why this study complements the cross-country analysis with two focused case studies, reported in Chapter 5 and 6.

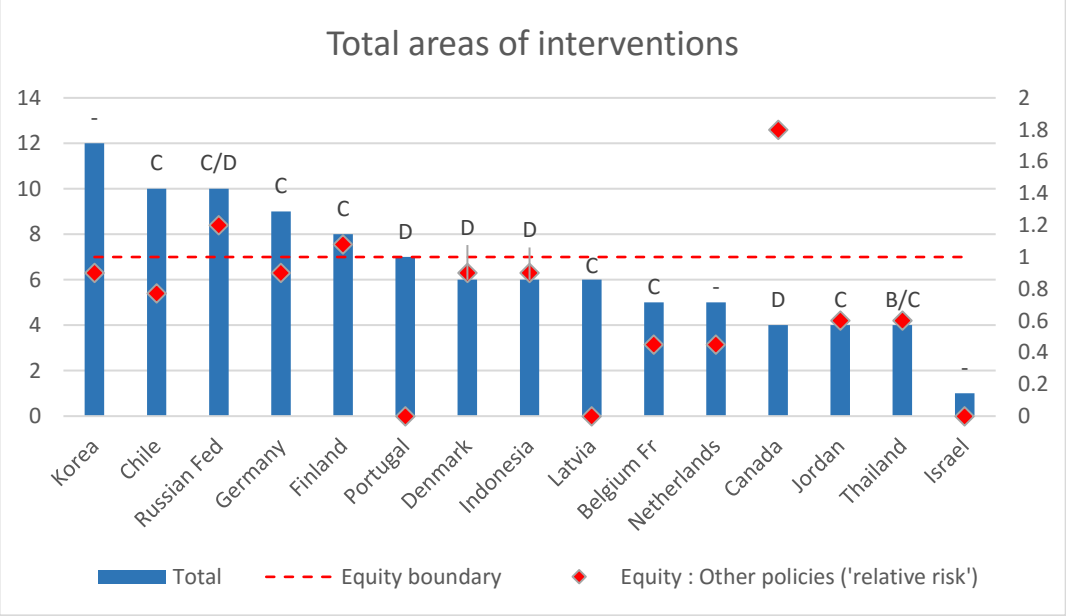
Finally, it is remarkable that countries that are routinely used by the OECD as evidence of the usefulness of the EPA were deemed by experts to have been only moderately affected by OECD recommendations. Brazil was mentioned in Chapter 3 because the claim that it 'offers a good example of how low-performing countries can use international benchmarking to improve their education systems' (OECD, 2010d, p. 102) was not matched by evidence; answers from the questionnaire add to that critique. Japanese efforts 'to promote more relevant learning' may well have translated into 'good results on the PISA test' (OECD, 2014f, p. 125) but, according to the expert's response, curricular relevance was not necessarily aligned with OECD recommendations. Despite the 'PISA shock', policy alignment in Germany has remained modest, and partial alignment can be found also in neo-liberal Chile or "model student" Finland. Portugal, however, is presented as a successful reformer by the OECD and the questionnaire supports this view.

Breakdown by country

Respondents were sent a second email linking an online questionnaire called the OECD *Policy Advice Implementation Questionnaire* (PAIQ; Aloisi, 2014; also Appendix 3, Table 2.D). The questionnaire asked respondents to select the specific policy areas in the EPA (Table 3.2) that had been translated into policy. Selected areas were coded "1" and non-selected areas were coded "0". Sixteen country breakdowns were collected in this fashion, which resulted in a 16 x 14 matrix of 0s and 1s (Appendix, Table 4.N). The Russian Federation features twice in the table: the two experts disagreed on the number of areas Russian interventions covered, but they agreed that four dealt with increasing time in education, reducing segregation for at-risk students, strengthening the teaching profession and standardising the educational outcomes.

Assuming Russian policies affected ten rather than four policy areas, country results can be displayed in Figure 4.9.

Figure 4.9: Number of recommendations in the EPA covered by country policies



Source: own elaboration using data from the expert questionnaire and the PAIQ.

The blue bars and the values on the left *y*-axis show the number of recommendations in the EPA that had been integrated within national policies, ranging from 12 (Korea) to 1 (Israel) with an average of about 6 per country. There is no evident relationship between the overall rating of country alignment with the EPA from the first experts' questionnaire (shown by the letters above each bar) and the number of areas affected by policy interventions. Given that, in most cases, both data were provided by the same expert, this suggests a certain variability, once again, in the individual interpretation of options such as 'to a modest extent' and 'quite a lot'. Not all columns are labelled because some experts only answered the PAIQ.

In terms of policy areas, 80% of countries in the sample have implemented or planned reforms to strengthen the teaching profession, 67% have worked on time spent in education and on the standardisation of outcomes and 60% on interventions targeting low achievers (Appendix 3, Table 4.N). On the contrary, reducing stratification and competition were the least popular policy options. This leads to the second part of the graph, represented by the red dots and dashed line elements.

The OECD recommendations from the EPA summarised in

Table 3.2 can be assigned to two groups, one comprising five recommendations on equity and one comprising the remaining nine. The countries' preference for equity policies over others can be expressed numerically in terms of relative risk, which is simply the ratio between the frequency of adoption of equity policies over the frequency of adoption of the other policies. This ratio is represented by the dots and the values on the right *y*-axis in Figure 4.9. The dashed line indicates the boundary ratio of 1: above the line, equity policies are given higher priority than all others combined, whereas below the boundary they are given lower priority.

For instance, reforms in Finland covered eight recommendations, three from the equity group and five from the other areas. In total, Finnish reforms addressed fewer equity recommendations, but policymakers also had fewer choices. If one considers instead that three *out of five* (60%) possible Equity categories were covered by Finnish reforms, against five *out of nine* (56%) of the remaining categories, it emerges that in fact policymakers in Finland adopted, within the available options, more recommendations from the equity area than from all other areas combined. In terms of relative risk, three out of five divided by five out of nine equals 1.08, which means that equity policies are slightly more likely to have been chosen than other policies.

There is a tendency for countries that have covered more policy options to have a higher ratio (though theoretically the two values should be unrelated), and the ratio is above the boundary in only three cases. Perhaps, equity recommendations are followed once other areas have already been addressed. Canada is an outlier: despite the few EPA sections touched by national reforms, half of them were about equity, specifically about reducing the gaps between genders and between the majority of students and minority groups.

At the other extreme, the value for Portugal is not fully consistent with OECD publications. The OECD reported that inequality increased between 2006 and 2012 and student dropout rate is still high, which may have led the expert to reckon that not enough had been done in terms of equity. Moreover, many Portuguese policies focused on other areas like 'training teachers, granting more autonomy and expanding assessments' (OECD, 2013f, p. 105). At the same time, however, Portugal did show some commitment to equity policies such as extending the end of compulsory education or the 'School Social Action' programme, whereby 'additional support, such as meals and books, is provided to disadvantaged students' (p. 104).

These results are a reminder of the limitations of this exercise. Responses were collected from a small sample of experts and, with the exception of the Russian Federation, only one set of answers per country are available. After analysing and comparing experts' responses from the initial questionnaire and the PAIQ, it is clear that the validity and reliability of the

two instruments could be improved. The efforts made to increase participation had to face the limited resources available. Recall also the difficulties in framing the EPA, whereby even the most thorough attempt (Table 3.2) has a degree of arbitrariness. For instance, it could be argued that Recommendation 4.1 on helping low-performing students could belong to the equity area, because disadvantage and low achievement are correlated, and similar arguments could be made regarding pre-primary education or resource allocation. A different categorisation would have produced different ratios.

The relationship between policy alignment and country outcomes

An ideal follow-up to the PAIQ would use expert answers (Korea, 12; Chile, 10; etc.) as evidence of policy alignment with OECD recommendations, and this evidence could be added as a time-invariant variable in a multilevel model of performance and equity trends. This was not the approach used in this research, however, for theoretical and practical reasons. On a practical level, the sample size was too small to warrant the use of multilevel modelling. On a theoretical level, values from the PAIQ are better placed on an ordinal scale than on an interval or a ratio scale; it makes no sense to claim that Korea is twice more aligned with OECD recommendations than Denmark, because the expert rating does not capture the number, strength and effectiveness of national interventions. For these reasons, the Spearman's rank correlation coefficient (Spearman's ρ) was deemed to be a more appropriate measure to capture the relationship between policy alignment and country outcomes.

PAIQ values of policy alignment were ranked (ties were allowed). To account for the countries' differential ability to translate policy plans into action, PAIQ ranks were weighted by country rankings in governmental effectiveness, one of the variables in the World Bank's World Governance Indicators database (see also Kaufmann et al., 2010). The weighting of PAIQ ranks by effectiveness ranks produced the index of policy alignment with OECD recommendations, or "EPA Index".

The EPA Index attempted to capture both the degree of policy alignment and its strength, producing different country rankings from those obtained using PAIQ values only (Appendix 3, Table 4.O). For instance, reforms in the Netherlands or Denmark covered few EPA categories, but if one takes into account that those few categories were probably linked to more effective policies than the same categories in other countries, then the Netherlands and Denmark are in a stronger position compared to Portugal or the Russian Federation, even though these latter scored higher in the PAIQ.

Table 4.9 shows the Spearman's ρ between EPA Index and country measures of achievement and equity. The row "Scores" is country performance, whereas "Inequality" is given by the Strength of the socio-economic gradient. "Level in 2012" refers to performance

and equity values in 2012. "Change first–last" is the change in performance and equity between the first and last PISA participation of this small country subset (The Netherlands did not participate to PISA 2000, Israel and Chile to PISA 2003, and Jordan to PISA 2000 and 2003). For the annualised change, performance values were taken from the PISA 2012 database, whereas values for equity were calculated via individual country OLS regressions.

Table 4.9: Spearman's rank correlation coefficient between PISA outcomes and the EPA Index of 15 countries

	Level in 2012			**Change first–last**			**Annualised change**		
	Read	*Maths*	*Science*	*Read*	*Maths*	*Science*	*Read*	*Maths*	*Science*
Scores	0.71***	0.68***	0.72***	-0.25	-0.31	-0.12	-0.27	-0.22	-0.11
Inequality	0.36	0.28	0.23	0.05	-0.33	0.15	0.03	-0.30	0.14

*** $p < 0.01$. *Source*: own elaboration using PISA, World Bank and the original equity databases, as well as data from the PAIQ.

Country performance ranks in 2012 and EPA Index are strongly positively correlated, but weaker relationships can also be seen between EPA Index and inequality (the Strength of the socio-economic gradient in 2012). In other words, the countries with higher policy alignment tend to be both the highest performers and, to a lesser extent, the most inequitable economies of the sample. The relationship between EPA alignment and performance changes is weak and negative, which captures the fact that Finland and Denmark have seen their performances decrease whereas Israel, Indonesia and Thailand improved. No clear relationship can be identified between the EPA Index and trends in equity.

Notice that even though the correlation between EPA alignment and country performance in 2012 is statistically significant, its size is driven by the *Governmental Effectiveness* variable. If one removes the interaction between the effectiveness and PAIQ values, ρ(PAIQ–PISA) is between 0.21 and 0.29 and is non-significant, whereas ρ(Effectiveness–PISA) is 0.66–0.71 and is significant. Therefore, these results provide only mixed evidence that EPA alignment is in any way related to PISA outcomes.

Some limitations of PISA and their influence on the trends of OECD and partner countries

Many times in this chapter, a disclaimer has had to be made that observed changes in performance and equity might be a statistical artefact. This is because there is strong evidence that PISA scores are affected by multiple sources of error, including construct validity, cultural and translation biases, sampling, scaling and equating, and other validity and reliability issues (summarised, for instance, in Hopmann, Brinek, & Retzl, 2007; Prenzel, Kobarg, Schöps, &

Rönnebeck, 2013). In this section, validation issues in PISA are conceptually organised in two categories: comparability across countries and comparability over time.

Comparability across countries

Sampling

A first issue with sampling in PISA concerns unit replacement when some schools decline to participate in the assessment. In theory, a country is considered to have an acceptable school response rate when at least 85% of the schools initially invited to participate do so; if the response rate falls between 65% and 85% (an 'intermediate' response rate) replacements can be accepted, whereas values below 65% are not acceptable (OECD, 2014h). If replacements are used, the 'acceptable' response rate threshold is raised to 95% comprising first and second respondents.

The OECD does not state what happens if response rates are below 95% after replacement, but even if it did, it is clear that these stringent rules are not adopted in practice. Countries are routinely given second and third chances regardless of their response rate. Influential economies such as the US, the UK and the Netherlands were allowed to take the test even when initial response rates were—on paper—not acceptable (Adams & Wu, 2002). Surely, sometimes response rates are so low that countries are excluded *ex post facto*, but this is an exception rather than the norm.

It is certainly good practice to be flexible but, as the late Sigbert Prais poignantly remarked, 'what is the point of stipulating a "standard", if it is at once replaced by something else?' (Prais, 2003, p. 151). When one considers that student scores are used for estimation when the student participation rate within a school (original or replacement) is as low as 25% (OECD, 2014h), it is evident that non-response bias is an issue. This is what Prais (2003, 2004) pointed out in an exchange with Raymond Adams (2003) with particular reference to England. Prais (2003) noted that the way the OECD deals with response rate was at best simplistic and that, considering effective participation at the school and student level, results for England may have been artificially raised 'by some 38 points' (p. 152).

A second issue with sampling concerns the representativeness of the PISA population. Student population coverage has been over 95% for OECD and non-OECD countries since the first administration, meaning that the PISA sample reflects the contexts and characteristics of most students in a country (provided all students in the sample take the test). The PISA sample, however, does not include students not in education at the age of 15. Therefore, the extent to which it represents all 15-year-olds in a country, including those who were in education and left, varies. With the progressive opening of PISA to partner economies, the number of countries whose PISA sample represents less than 80% of the 15-year-old

population has increased from 5 to 17 in 2012. The 20% or more of young people not covered by the PISA sampling frame effectively represents data missing not at random, but it is unclear how this may influence country outcomes.

Some strong improvers such as Brazil, Indonesia, Mexico and Turkey have increased their 15-year-old population coverage (Turkey moved from 36% in 2000 to 68% in 2012). The OECD interprets coverage increase as a sign that countries were more effective in reaching out to young people that would not have entered education otherwise (see for instance Box I.2.4 in OECD, 2014g). As the OECD argument goes, if these newly-enrolled students came from the bottom of the performance distribution, then Brazil, Indonesia, Mexico and Turkey managed to increase their performance in spite of the negative effect on their average score of a greater number of lower-achieving students, which is an excellent result. However, it is not known whether increases in the coverage index capture increases in "incidence" of enrolment rather than different ways to measure it.

While it is 'a critical component of the sampling frames', in fact, student enrolment is estimated on data that are 'rarely available at the time of school sampling' (OECD, 2014h, p. 71). In Mexico there were fewer 15-year-olds in 2012 than in 2000, yet the enrolled population was estimated to be 34% larger (Appendix A in OECD & UNESCO-UIS, 2003a; OECD, 2014g). Quality controls are in place and it is likely that, in this case, 'large deviations between the total national number of 15-year-olds and the enrolled number of 15-year-olds were questioned' by the PISA Consortium (OECD, 2014h, p. 81), but the OECD does not release information about the investigation procedure.

The school sampling frame is another critical estimate, 'NPMs were therefore advised to be diligent and thorough' in its construction (OECD, 2014h, p. 70). The OECD concedes that 'the definition of a "school" is difficult' (p. 85) and that some countries choose other sampling units instead. In Mexico, in 2003, 'schools where instruction is delivered in shifts were split into the corresponding units' (OECD, 2004b, p. 328). According to the international and technical reports, this was not the case in 2009 or 2012 and it is unclear what choice was made in the intermediate cycles. One has to accept that 'differences were queried' (OECD, 2014h, p. 83).

Finally on this point, three additional factors may have affected the scores of countries in the southern hemisphere: the international target population is conceived 'to better fit the age structure of most of the Northern Hemisphere countries' (OECD, 2014h, p. 66); the end of the testing window has moved from 31 October to 31 August between PISA 2000 and 2003; and the 'no testing period' shrank from the first three months to the first six weeks of school between 2003 and 2006 (Adams & Wu, 2002; OECD, 2005b, 2009c).

Translation and item-by-country interaction

Grisay et al. (2007) found evidence of regional clusters in country response patterns to PISA items. They remarked that the PISA 2006 booklets 'did not function exactly in the same manner in all participating countries, and that the differences in languages used probably played a role in the item/country interactions' (p. 253). They also found lower equivalence between the international version of the assessment and national versions in low-GDP, Middle Eastern and non-OECD Asian countries, suggesting that results from those regions were less comparable than results from Western OECD countries. Similar issues were reported for all three PISA domains between 2000 and 2006 (Akour, Sabah, & Hammouri, 2015; Grisay, Gonzalez, & Monseur, 2009; Le, 2009; Yildirim & Berberoĝlu, 2009).

In this regard, Kreiner & Christensen (2014) critiqued the application of the Rasch model[15] to PISA on the grounds that it fails to account for the strong differential item functioning [DIF] that can be observed in all PISA items and booklets. Existence of DIF means that item difficulty varies systematically across countries when compared with the rest of the scale; since mean scores are estimated assuming that there are no such variations, results for countries in which the DIF is stronger are artificially inflated or deflated. Schleicher (2013d) fiercely rejected Kreiner & Christensen's (2014) argument, but there is no doubt that the items selected for each PISA cycle can and do affect country scores in ways that are not related to student ability. This was also pointed out by Goldstein (2004), who argued that DIF 'essentially precludes any comparison based upon a single scale' (p. 328); and by Wu (2010): 'Given that the inclusion of particular […] units and items is largely by chance […], we query about the level of confidence we can place on the results' (pp. 21–22).

Part of the problem lies in the philosophy of the Rasch model. Rasch is not a statistical model that can be fitted to the data; it is a measuring approach requiring the data to fit the model. For Rasch to work, test development must be conceived as an iterative process whereby 'misfitting' items are followed up after each assessment and amended to better fit the model requirements (unidimensionality, local item independence etc., see Bond & Fox, 2001). This approach to measurement requires procedures that are hardly compatible with the times and scale of PISA, yet the model is applied with 'almost no reference to the debates about […its] appropriateness' (Goldstein, 2004, p. 328) and in spite of evidence suggesting that models relaxing Rasch assumptions could prove more useful for the purposes of the assessment (Oliveri & von Davier, 2011).

[15] The Rasch model (Rasch, 1980), in its original form, assumes that the probability that a candidate will answer an item correctly depends exclusively on the candidate's ability and the item's difficulty. For more information, http://www.rasch.org/rasch.htm (Accessed 16 April 2016)

Comparability over time

Problems in the application of the Rasch model

Issues of DIF and other violations in the assumptions of the Rasch model also have consequences on the comparability of country outcomes over time because of their effect on link items. "Link items" are items that are common to all assessments, although the exact number changes by domain and by cycle, because the item pool is progressively renewed (Appendix 3, Table 4.P). They are used to equate the PISA scales from one cycle to the next so that assessments from different years can be compared, but this practice is affected by two main issues.

Firstly, PISA tests are chain-linked: PISA 2012 is anchored to PISA 2009, which is anchored to PISA 2006 and so on (OECD, 2014h). This means that country outcomes could experience the same 'drifting' of student scores in some national assessments (Tymms, 2004), whereby achievement increases as a result of errors being brought forward from one year to the next. The size of these errors does not depend on the student sample but on the much smaller sample of link items, which makes linking error the greatest source of variability in country-level trends (Haertel, 2004; Michaelides & Haertel, 2014). This could be a problem especially for Reading, the domain with the lowest number of link items (Appendix 3, Table 4.P).

Secondly, the linking error in PISA changes by subject and by assessment cycle but is considered to be the same for all countries. If link items have DIF, however, additional error will be introduced because of the item-by-country interactions. It has been shown that if the linking error took DIF into account, the 24-point drop of Japan in PISA 2003 (the largest decrease at the time) should be considered statistically no different to zero (Monseur & Berezner, 2007; Wu, 2010).

More generally, Gebhardt & Adams (2007) demonstrated that 25% of observed changes between 2000 and 2003 were due to DIF and Wu (2010) argued that linking errors alone could account for score differences of up to 40 points. Considering that linking errors are but one source of variability, and that the greatest variability tends to affect lower-performing countries and Reading scores (G. Brown, Micklewright, Schnepf, & Waldmann, 2007; Monseur, Baye, Lafontaine, & Quittre, 2011), it is not unreasonable to think that many of the country fluctuations visible in Figure 4.1 might be, from a statistical perspective, the same as flat lines.

OECD reactions to the limitations of PISA

After enumerating issues with the sampling, modelling and interpretation of PISA (many of which were described above), Wuttke (2007) concludes that 'PISA is state-of-the-art' but is also affected by 'a plethora of inaccuracies' (p. 261) compromising by an unknown degree the accuracy of the reported scores and the validity of the statistical procedures.

When the PGB decided that 'the establishment of reliable trends should become the overriding priority', it commissioned an external review of the PISA design (Mazzeo & von Davier, 2008, p. 4). An exchange between the reviewers and the Consortium followed: the Consortium critiqued the reviewers' inability to demonstrate the effectiveness of some proposed changes, whereas Mazzeo & von Davier rebutted, similarly to Wuttke (2007), that the complexity of PISA paired with the scarcity of available data made it impossible for anyone to go beyond broad recommendations (Mazzeo & von Davier, 2008). There is a feeling that when the review took place it was already too late: to use a metaphor, the PISA machine had been assembled by so many hands that by then no individual knew how the components interacted, and any intervention was precluded by the fact that the engine was already running at full speed.

The OECD has recently started to acknowledge that sometimes the specific mechanisms driving PISA trends are 'unclear', that existing explanations are incomplete and that 'other reasons should be explored in the future' (OECD, 2014h, p. 272). It noted that 'the complexity of the PISA 2000 design is such that the impact of this on the item parameter estimation and hence the equating is unclear' (OECD, 2012e, pp. 216, 237). But for some researchers, recognising that there are issues is unsatisfactory if not followed by action, especially in the sense of greater disclosure of assessment data (Hopmann et al., 2007).

Hopmann & Brinek (2007) lament that the behaviour of the PISA consortium in Germany towards researchers resembled that of pharmaceutical companies facing a potential scandal: firstly, answering criticism with silence; 'if that is not enough, the next step is often to raise doubts about the motives and the abilities of those who are critical of the enterprise'; then, 'acknowledge some problems, but […] insist that they are very limited in nature and scope, not affecting the overall picture'; and finally, downplay the novelty of the issues and suggest that they have already been dealt with, often by referencing some 'opaque technical reports' or 'unpublished papers' (pp. 14–15).

Of course, comparing PISA with ill-conceived drugs is probably 'overreaching' (Hopmann & Brinek, 2007, p. 15), and it should be appreciated that the OECD has officially displayed an interest in improving the quality of PISA. The external review commissioned to Mazzeo & von Davier (2008) is one example; another is the change in the computation of the linking error after the critique of Monseur & Berezner (2007). At the same time, the many

voices inside the organisation contribute to creating a sense of friction between the scientific and the political side of the endeavour.

At a meeting in which this author was present, a participant challenged Schleicher on the fact that statistically-inflated score changes, such as the Japanese drop between 2000 and 2003, had real consequences on national policies (on this matter, Takayama, 2008). Schleicher commented that countries should be more patient and wait for more consistent trends to emerge before embarking in systemic reforms. He added that international reports consistently remind readers of the correlational nature of the findings. This is true, yet the overarching message is another: 'even relatively small differences between countries in the average performance of students, where they are statistically significant, should not be overlooked' (OECD, 2007b, p. 115).

Conclusions to Chapter 4

This chapter has collected and analysed large-scale evidence on the use of the EPA for national policymaking and its plausible effects on country outcomes. Firstly, educational outcomes in performance and equity were analysed using PISA data as well as data from an original dataset of equity measures. Secondly, the different "effects" of changes over time in three country-level factors (economic productivity, equitability and access to pre-primary education) on PISA trends were explored.

It was expected—and confirmed with the analysis—that the relationship between output and outcome would be stronger for the factor that is more malleable by specific education policies (access to pre-primary education), smaller for the "generic" educational factor (equitability) and even smaller for the non-educational one (productivity). This helped to set a first empirical limit to the extent to which the "effect" of education policies on PISA may be detected: about 0.3–0.4 PISA points per percentage point increase in a relevant indicator (in this case, gross enrolment ratio). It also confirmed the validity of OECD recommendations when it comes to early childhood policies, though it seems that most countries were already moving in that direction.

A third step linked policy alignment to outcomes, by collecting data from educational experts from countries participating to PISA through an email questionnaire and a follow-up online survey. Using their responses and a World Bank indicator of governmental effectiveness, an index of country implementation of OECD recommendations—the EPA Index—was produced and correlated to PISA scores. No clear relationship emerged between implementation of the EPA and changes in performance or equity.

These analyses were followed by a brief validation of PISA, which served to moderate the strength of the inferences that could be drawn from the three steps. A summary of the evidence

gathered to answer the research questions 2 and 3 follows, but generally, the use of country scores or equity measures to justify policy stances seems unwarranted. This is primarily because of the large error associated with individual country outcomes and, to a lesser extent, because of the risk that actual changes may be driven by factors which are non-malleable by education interventions. There is little evidence that the EPA is being followed or is effective, apart for one recommendation (access to pre-primary education) which is strongly backed by the literature and may indeed be associated with country improvement.

Evidence of country alignment with OECD recommendations (RQ2)

Despite the wide scope of the chapter, some inferences about country alignment with OECD recommendations are possible.

Firstly, if there were widespread interest in following the OECD advice of increasing within-country equality, this is not evident from the data. There is some evidence that inequality is decreasing, but the effect concerns only OECD countries and can only be detected in the correlation between the Strength of the socio-economic gradient and Reading scores. Furthermore, the rate of decrease is very small (3 percentage point in 12 years) and in general equity trends are very stable.

On a specific policy level, there is evidence of policy convergence in pre-primary enrolment over time, but it is difficult to identify global shocks after 2000. Gross enrolment ratios have increased by about 1.2 percentage points per year since 1988 in both OECD and partner countries. The rate of increase is accelerating, but by a very small amount which does not substantively affect average estimates. It was argued in Chapter 3 that pre-primary education has never become a priority for the PGB despite its relevance for student trajectories. Still, the possibility that PISA results have influenced its rate of expansion in participating countries should not be discarded. Some individual country increases seem particularly sharp, especially after 2008. These results could be followed up in future research.

Finally, many experts stated that their countries were more in line than not with the EPA, but a follow-up questionnaire (the PAIQ) revealed that, in a small sample of 15 countries, only five had implemented policies covering more than half of the EPA and only three had privileged equity policies. Given that answering the PAIQ required more time and commitment, it is likely that responses more accurately reflected country contexts than those from the first expert questionnaire. These findings add to the evidence that there is a gap between the use of PISA, the acknowledgement of OECD values and actual country alignment with the EPA.

Evidence of impact of EPA-aligned policy (RQ3)

In terms of performance, trends are highly linear and within-country variance accounts for 2–5% of the total. On a global level, countries have improved on average by 0.6–1.1 point per year depending on the domain, and most score changes are within ±20 points regardless of the distance in time between the two measurements. Altogether, these results suggest that any policy effect, whether aligned with the EPA or not, must be very small. It probably takes four or five PISA administrations to claim with enough confidence that country scores did indeed change.

Larger changes were found among partner countries and for Reading scores, but the extent to which they had been brought about by education reforms and had not been inflated by statistical error was called into question. There was also some evidence that global trends in Reading and certain country trends may be positive and curvilinear: if these findings were confirmed in future administrations, they could provide indirect evidence of the impact of the education reforms of the early 2000s on PISA performance (though not of policy alignment).

In terms of equity, trends are generally stable; half of the countries saw no change and the remaining half is equally split between countries in which inequality is increasing and countries in which it is decreasing. Trends in equity and achievement do not seem to be strongly related, either: OECD and partner countries have similar levels of equity, yet their results are very different; between-country variance is decreasing in performance and increasing in equity; and findings from multilevel models have shown that equity levels are unrelated with PISA outcomes. Overall, if countries have been implementing equity measures following the EPA—and this does not appear to be a priority, according to the PAIQ—these are having little effect on a global scale.

There is also some evidence that both performance and equity are decreasing in higher-achieving countries such as Korea, Finland and New Zealand, whereas in Germany, Mexico and Turkey they are both improving. In fact, a moderate negative correlation between changes in inequality and changes in achievement was found. More data from future PISA cycles could be collected to establish whether the observed relationship was the result of sampling error or not and what role policy alignment with the EPA might have had.

GDP trends were not related to detectable changes in scores, whereas trends in participation in early childhood education were, albeit weakly: countries that increased access to pre-primary education when the PISA cohorts could benefit from it experienced small PISA score increases (but only in Reading). These findings provide some reassurance that certain EPA-aligned education policies can be effective and that their effects might be captured by PISA.

At the same time, the role of societal trends outside the control of MoEs should not be underestimated. At the student level, the literature agrees that family background is a reliable predictors of later outcomes, which means that socio-economic policies may matter at least as much as education reforms. At the country level, a threshold was identified in the relationship between country wealth, expenditure per student and PISA outcomes: under certain levels, the association between economic factors and country scores becomes so strong that it is hard to deny that "money matters" for educational outcomes in poorer countries.

Unfortunately, expert responses were too few to provide reliable data on the influence of EPA-alignment on achievement or equity. A weak positive relationship was found between expert responses in the PAIQ and ranking in PISA, and the relationship was stronger when a measure of policy alignment—the EPA Index—was used; in this latter case, however, the association was probably driven by an interacting variable (governmental effectiveness). The correlation between EPA alignment and performance changes was weak and negative, capturing the fact that Finland and Denmark saw their performances decrease whereas Israel, Indonesia and Thailand improved. A weak positive correlation instead emerged for country levels of inequality in 2012, but not for equity trends.

Altogether, these results provide little support to the claim that EPA alignment is related to PISA outcomes, at least on a very large scale. This is mostly due to the small effect sizes paired with lack of data, and therefore this impression might change with future cycles. Case studies follow in Chapter 5 and 6 to investigate the issue on a more local level.

Chapter 5: The use and impact of OECD education policy advice in Ireland

The cross-country analysis in Chapter 4 was complemented by study visits to Ireland and the French Community of Belgium. A comparative outline of the two education systems is provided in the Appendix. The analyses in this and the next chapter were guided by the six case study questions (CSQs) listed in Chapter 2. The chapter is articulated in the following parts:

- A historical overview of the main educational issues in the country before its first PISA participation, including country relations with the OECD, participation in other LSAs and the key educational legislative framework.
- A summary of the targeted EPA received after each PISA cycle, drawing from the national version of the report.
- A detailed analysis of selected policies, covering the context surrounding the policy interventions, their degree of alignment with the EPA and, when possible, any evidence of implementation and impact.
- A summary of country trends in policy alignment with the EPA.
- An evaluation of the plausible impact of national reforms on PISA outcomes.

The first step helps to frame PISA within an existing national context. It shows how receptive a country may have been to the new assessment and any pre-existing alignment with the EPA (CSQ4). The second step accounts for two issues: that a country may have received specific recommendations beyond the general policy advice summarised in Chapter 3, and that the interpretation of PISA outcomes by the national project managers and their policy advice might not be in line with that of the OECD (CSQ1). The third step helps to compare the experiences of both countries in similar areas whilst taking into consideration the complexity of policymaking and the nuances in policy alignment (CSQs 2–5). At the end of the chapter, the fourth and fifth steps organise the evidence collected to answer the overall research questions 2 (country alignment with OECD recommendations) and 3 (impact of EPA-aligned policies).

In the case of Ireland, it is argued that the country has always been receptive to OECD recommendations but since 2005 policy alignment has increased even further. There are few policies satisfying the criteria of being aligned with the EPA, having been implemented effectively, and being sufficiently old, that effects on student learning can be evaluated; however, DEIS and Project Maths provide evidence that EPA-aligned policies can be effective in raising equity and performance.

Historical overview

The evolution of education between religion and market economy

The 1937 Constitution established that the rights and responsibilities for children's education fell upon parents; the only obligation of the state was to arrange for (but not necessarily supply) free primary education. As a result, Irish education has traditionally been publicly-funded but privately-managed, specifically by the Catholic Church. The non-interventionism of the state has been criticised as a 'pious stasis': for many years, the function of education was to reproduce 'a certain social type, pious, familial, loyal to the native acres, culturally ingrown and obedient to clerical guidance in matters moral and intellectual' (Garvin, 2004, p. 184; quoted in O'Connor, 2014, p. 197).

The 'triumphalist' period of the Catholic Church in education (O'Donoghue & Harford, 2011) lasted at least until the mid-1960s, when the Department of Education and Skills (DES) began to play a more active role in the development and implementation of education policies following 'the increased participation rates from the 1960s, the development of the comprehensive and community schools, and the expansion of the vocational sector to include Regional Technical Colleges (now Institutes of Technology)' (Eurydice, 2016c).

The turning point in the transition from a 'theocentric' to a 'mercantile' paradigm (O'Sullivan, 2005) was *Investment in Education* (Government of Ireland, 1965), a 'technical study of trends in Irish education and of the use of human and material resources in that system' (p. xxxiii). The report was prepared by a survey team appointed in 1962 and it benefitted from 'the guidance' of the OECD (p. xxxiv), which also contributed with technical and financial support. The advisory role of OECD 'gave the report a neutral identity' (O'Sullivan, 2005, p. 141).

The authors provided only one formal recommendation, but it was sufficient to change the role of the state in educational affairs: 'the creation in the Department of Education of an educational development unit' (Government of Ireland, 1965, p. 387) 'to continue the kind of work that [the] survey [had] begun' (p. 350). More implicitly, the report also 'indicated numerous reforms required in the national system of education' (OECD, 1991, p. 7).

Despite its pivotal contribution, *Investment in Education* acted 'as a means of extending and legitimating a paradigmatic shift that had already occurred' (O'Sullivan, 2005, p. 137)—a shift that had led Ireland to participate 'eagerly' (p. 136) in the OECD Washington Conference of 1961 and to be the first country to volunteer to undergo peer review.

The new paradigm owed much of its success to the support of Irish parents, who saw their role change from 'mute followers' to decision-makers: human capital theory not only provided them with 'an understanding of education that they shared but it also named and legitimated

in official discourse what they had previously lived namelessly' (pp. 154–155). Since then, Irish parents have continued to request greater empowerment, sometimes advocating accountability measures opposed by teachers, such as being able to access achievement data or school performance reports (Kellaghan, McGee, Millar, & Perkins, 2004).

The emergence of human capital theory was accompanied, but not matched, by discourses on equity, which captured the interest of intellectuals and researchers thanks to the Coleman Report (J. Coleman et al., 1966) but remained at the outskirts of the political arena. Even the introduction of comprehensive schooling was framed as a practical need rather than a moral imperative. An exception was gender equality, appearing in the 1970s as a highly systematised discourse 'incorporating an unambiguous belief system, explicit apparatus, together with instruments and mechanisms for the articulation of correct practices, the monitoring of behaviour and action, the arbitration of doubt and dissent, and for its reproduction over time' (O'Sullivan, 2005, p. 425).

Equity became a valuable goal of education only in the 1980s, though it was still interpreted in economic terms. In agreement with the OECD view of the time (and to some extent today), the Commission on Taxation wrote that 'education policy may be perhaps the most powerful weapon available to democratic government in seeking to introduce equality in the distribution of incomes' (quoted in Fitzgerald, 2001, p. 156). Once it entered the system, addressing educational disadvantage became a staple of the educational agendas of both left and right-wing ministers[16] (see for example MoE O'Rourke's statement of educational priorities in OECD, 1991).

O'Sullivan (2005) argues that discourses on equity caught on across the political spectrum because of their fuzziness, which made them applicable to a range of contrasting situations. The term used by the author is "pastiche"—a form characterised by 'avoidance of dissonance and contestation, deferral and dispersal in understanding disadvantage, the enjoining of agreement, the mobilisation of diverse participants, and the orientation and facilitation of action' (pp. 322–323). This recalls the multi-vocal nature of OECD indicators mentioned in Chapters 1 and 3.

Many reforms have taken place since the 1980s, yet it has been argued that 'the structure of Irish second-level education remains largely as it was in 1965' (O'Connor, 2014, p. 204): primary and secondary-level education is segregated on grounds of religion and gender, and 'the state's commitment to education still does not extend much beyond the payment of salaries' (p. 203). Some of the interventions reported later in this section suggest that the role

[16] A list of education ministries is available at http://www.education.ie/en/The-Department/Ministers/Ministers-for-Education-From-1921-to-Date.html (Accessed 16 April 2016)

of the state is more incisive than claimed by O'Connor, but they provide little evidence of changes in religious and gender segregation.

International relations: the OECD and the European Community

Three factors contributed to shaping Irish education: 'the international restructuring of economies, the impact of reports and recommendations from the European Community (EC) and the influence of the OECD' (Drudy, 2000; cited in Halton, 2003, p. 334). Connections between the OECD and Irish policymakers have existed since the establishment of the organisation in 1961. The 1965 *Investment in Education* report was prepared under the aegis of the OECD, though some claim that its role was much more limited than commonly assumed (Hyland, 2014). In 1969, a follow-up report on the state of the education system highlighted the following weaknesses: the inefficiency of small primary schools; traditional teaching methods and low attention to physical sciences, biology and mathematics; teacher training; and the fact that education in Ireland was a conservative system based on the transmission of values and reflecting 'the [traditional] social structure of the community […] even when the circumstances of the community have undergone important changes to which education must respond' (OECD, 1969, p. 84).

In 1973, Ireland joined the Economic European Community and 'adopted a strong pro-European stance influenced by its desire for still closer links with mainland Europe' (OECD, 1991, p. 11). When the OECD inspectors returned, in 1991, they praised increased access and better curricular experiences for pupils in primary education (thanks to the well-received *Curaclam na Bunscoile* [primary school curriculum] of 1971) but they also reported that 'the system as such [had] remained largely the same' as 20 years earlier (OECD, 1991, p. 36). Systemic change 'was not planned methodically but expanded in piecemeal fashion in order to respond to importunate pressures' (*ibid*.). They found an innate conservativism, made worse by the presence of powerful interest groups outside the government and the absence 'of a purposeful central authority having the political will, administrative capacity, and requisite financial resources to formulate and implement reforms' (*ibid*.). The OECD also highlighted the lack of school and pupil data, the lack of consistency in teacher training and education, and the lack of a clear role for the Inspectorate: where primary school inspectors actively evaluated pupil and teacher performance, secondary school inspectors rarely visited schools and used most of their resources administering inadequate national examinations.

The Minister replied to the inspectors that 'the Irish education system and the results that it produced compared favourably with those of any country' (MoE O'Rourke in OECD, 1991, p. 116), but many of the OECD recommendations eventually flowed into the new Green Paper

for education (DES, 1992; Halton, 2003). Consequences of the OECD report on Irish policymaking can be seen for instance in the following excerpt from a debate in the *Seanad*[17]:

> 'The 1992 OECD report referred to in the Green Paper indicated that, while the problems of basic literacy were relatively low, the problems of functional literacy were alarmingly high [… The] Department has commissioned the Education Research Centre […] to carry out a national survey on literacy. This survey will be part of an international one led by those who produced the OECD report to which I have referred.' (MoE Bhreathnach, 1993)

The Green Paper, 'which was regarded as an attempt to introduce an "enterprise culture" into schooling' (Halton, 2003, p. 332), was followed by an unprecedented mobilisation of stakeholders in education that found its highest expression in the National Education Convention of 1993. Proceedings from the Convention (Convention Secretariat, 1994) informed the White Paper (DES, 1995) and, to some extent, the main legislative effort of that period: the Education Act, 1998 (Oireachtas, 1998). In all this, the OECD never quite disappeared from the picture: 'What we are doing in Ireland is of considerable interest to our partners, and the OECD is specifically monitoring this process where the partners in education are being consulted, heard and responded to before the "tablets of stone" [the government's White Paper] are finished' (MoE Bhreathnach, 1994).

Key educational framework

The Education Act (1998) was the first comprehensive legal framework for education since the foundation of the state. It sets the educational goals for the country (§ 6), the objectives of schools (§ 9), the role of the Inspectorate (§ 13), of the school Board of Management (§ 15) and of the School Principal (§ 23). It also establishes institutions such as the Educational Disadvantage Committee (§ 32) and the National Council for Curriculum and Assessment (NCCA, § 41).

Table 5.1 shows the remarkable alignment between the Education Act and the EPA both in contents and language, particularly in matters of equity. It is plausible that the views of the OECD and the European Commission, which had already entered the Irish policy discourse through the Green and the White Papers, contributed to shaping the founding Act of Irish education.

Table 5.1: contact points between the Education Act, 1998, and the OECD education policy advice

Educational time
▪ § 25(c) and 30(1)(c) grant the MoE the authority to prescribe—following consultation—'the length of the school year, the school week or the school day' and 'the amount of instruction time to be allotted to each subject'

[17] The Parliament of Ireland (*Oireachtas*) is divided into two Houses: The House of Representatives (*Dáil Éireann*) and the Senate (*Seanad Éireann*). The Irish names are used in this document. All debates are available from http://oireachtasdebates.oireachtas.ie/ (accessed 16 April 2016).

Equality of opportunities to learn
▪ All actors should promote 'equality of access to and participation in education' § 6(c), 21(2), 41(2)(j), 41(3)(b). ▪ All actors should work toward the inclusion of children with a disability or other special education needs § 6(a), for instance by promoting 'effective liaison' of all who have 'a special interest in, or experience of' SEN education § 6(g). ▪ Schools should 'promote equality of opportunity for both male and female students' § 9(e) and the NCCA should 'promote equality […] of instruction in any particular subjects between male and female students' § 41(2)(j). ▪ "Educational disadvantage" is defined as 'the impediments to education arising from social or economic disadvantage which prevent students from deriving appropriate benefit from education in schools' § 32(9), an Educational Disadvantage Committee is established and resource allocation should account for 'the level of disadvantage of student in schools' § 12(1). ▪ Schools should 'ensure that students have access to appropriate guidance to assist them in their educational and career choices' § 9(c) and the curriculum for guidance and counselling may be set by the MoE § 30(1)(d).
Quality of provision
▪ All actors should 'promote best practice in teaching methods' § 6(f), and for this purpose the NCCA should 'from time to time to review the inservice [CPD] training needs of teachers' § 41(e). ▪ All actors should work to 'enhance the accountability of the education system' § 6(l). § 13 addresses the work of the Inspectorate but largely 'reflects the existing practices of the Inspector in statutory form' (McElduff & Sheehan, 2001, p. 5). Principals should 'regularly evaluate students and periodically report the results of the evaluation to the students and their parents' § 22(2)(b). ▪ Principal and teachers should 'collectively promote co-operation between the school and the community which it serves' § 22(2)(c).
Student performance
▪ § 30(1)(a) and (b) grant the MoE the authority to prescribe the school curriculum. ▪ The NCCA should devise 'mechanisms whereby students who have problems achieving their potential may be identified as early as practicable and assisted' § 41(b). ▪ The NCCA should devise methods to assess students 'having regard to national and international standards and good practice in relation to such assessment' § 41(d).

Not all provisions in the Act agree with the EPA. For instance, § 6(e) states that all actors in education should 'promote the right of parents to send their children to a school of the parents' choice', which is contrary the EPA and negatively impacts on the equity of the system.

Previous participation in international assessments

Before PISA, in 1990/1991 Ireland participated in both the IEA Reading Literacy Study (RLS) and the Educational Testing Service's (ETS) Second International Assessment of Educational Progress (IAEP2, Lapointe, Askew, & Mead, 1992; M. O. Martin, Hickey, & Murchan, 1992). Results from IAEP2 were collected in *Education at a Glance* (CERI, 1993), which in turn was the reference source for the Irish government to claim in the White Paper that 'the overall science achievement of thirteen-year-old students in Ireland is low compared to a number of other OECD countries' (DES, 1995, p. 23). O'Leary (2001) demonstrated, using data from both IAEP2 and TIMSS 1995, that the poor performance reflected sampling choices rather

than actual learning but, he argued, policy-makers continued to use IAEP2 results at face value. In 1994, Ireland took part in the OECD International Adult Literacy Survey (IALS; DES, 1997; see also Thorn, 2009, for a summary of the adult literacy studies).

In 1995, Ireland took part in TIMSS. Findings were reported over three years respectively focusing on grades 7-8 (Beaton, Martin, et al., 1996; Beaton, Mullis, et al., 1996), grades 3-4 (M. O. Martin et al., 1997; Mullis et al., 1997) and grades 11-12 (Mullis et al., 1998). Ireland emerged as an average-ranking country on most indicators, with the exception of mean performance and improvement between 3rd and 4th grade. The primary grades attracted most positive comments, whereas the small performance improvement in the transition from primary to secondary education was highlighted in more than one instance. Other negative points were teachers' beliefs about mathematics and assessment, student beliefs about science, the reliance of teachers on textbooks, time dedicated to the instruction of science and to homework, and Irish teachers' complaints about class sizes. This was the last TIMSS assessment until 2011. Most interviewees were reluctant to discuss why Ireland interrupted its participation, but there was some agreement that the decision was unexpected and was pushed by a senior advisor to the MoE—perhaps the Chief Inspector Eamon Stack, who was in charge between 1997 and 2009 and was replaced by Harold Hislop in 2010.

In 1998, the DES commissioned the Educational Research Centre (ERC) 'to conduct a national assessment of the mathematics achievement of pupils in fourth class in primary schools' (Shiel & Kelly, 2001, p. v). Similar low-stakes assessments had already taken place after the introduction of the *Curaclam na Bunscoile* of 1971, but the National Assessment of Mathematics Achievement 1999 was particularly relevant for three reasons: it took place 15 years after its most recent predecessor, it restarted a five-yearly practice that, since 2004, also includes the assessment of English reading, and it was statistically linked to TIMSS 1995 and 1999 (Shiel & Kelly, 2001).

Overall, Ireland's attention to international and standardised assessment pre-dates PISA and may have made its adoption easier. The first mention of PISA in a parliamentary debate dates to 22 October 1998, when the MoE Micheál Martin was asked whether the DES was involved in any study on literacy levels in secondary education:

> 'My Department is participating in an OECD project which aims to produce policy relevant and internationally comparable indicators of student achievement. The project is called PISA, Programme for International Student Achievement. The project spans nine years and will be concerned with three major literacy areas […]. This will be the first time when direct international comparison will be possible in three key curricular areas at 15 years of age, an age when the vast majority of pupils sit the junior certificate examination.' (M. Martin, 1998)

Thus, the policy relevance of PISA was established from the start because of its link with the Junior Certificate Examination (JCE) taken by the majority of 15/16 year-old students at the end of compulsory education. Notice that the salience of PISA was predicated on its ability to provide information on 'key curricular areas', not to assess skills for life. The Minister also thought that the "A" in "PISA" stood for "Achievement" instead of "Assessment". This is a minor but interesting mistake, as there might be policy implications when treating something as a strategy for achievement rather than an assessment. Similar mistakes were made in parliamentary debates in Belgium. They are indicative of a certain inattention to small but nontrivial aspects of large-scale assessments by policymakers that could easily translate into loss of alignment when international recommendations are adopted on a national level.

The education policy advice for Ireland

In Ireland, 'PISA is implemented by the Educational Research Centre on behalf of the Department of Education and Skills. A national advisory committee oversees the implementation of PISA and advises on all major aspects of the study, including reviewing the assessment materials and providing input into national reporting' (ERC, 2015). The ERC also produces teacher guides to PISA (for PISA 2000, see Cosgrove, Sofroniou, Kelly, & Shiel, 2003), in which PISA findings are presented in the context of country policies. These guides provide the same information as the national reports and they do not seem to be trying to affect classroom practice, as recommendations to teachers and schools are general and limited to a few paragraphs. Perhaps they are a form of 'opinion mobilization' (see Table 2.1: Levin's categorisation of policy levers), a strategy to ensure policy absorption and implementation. Because of their unclear function and impact, however, they are not considered any further in this chapter.

As mentioned above, after the release of the international report, a national version is prepared by the ERC. It can be shown (Appendix 3, Tables 5.A–D) that the OECD and the ERC generally come to the same conclusions with respect to Irish performance but this is not always the case. Sometimes, this is due to the more extensive analyses carried out by the ERC. Sometimes, however, OECD findings are absent from the ERC reports or the same data are interpreted differently.

This is a reminder that the national project managers do not simply reproduce OECD information, but are rather 'brokers' of knowledge (Meyer, 2010; Sriprakash & Mukhopadhyay, 2015). According to the principal–agent theory a research centre might, depending on its statute and leadership, acquire sufficient autonomy from its principal to add its own "voice" to that of the OECD. To understand the degree of autonomy of the ERC, this author tried to access further information about the its remit, mode of financing and

accountability, but without success. It is known that it was established in 1966 with a gentlemen's agreement between the Dean of St. Patrick's College and the founder, and that it has enjoyed a good degree of independence until very recently, when it has started to become more integrated with the MoE.

The mechanisms whereby an 'intermediary' organisation (Honig, 2004) like the ERC brokers the EPA to Irish policymakers were not the focus of the current study, but the fact that the centre's contribution is not limited to administering PISA and writing the national report should be taken into account. Differences between the OECD and the ERC goals and values may translate into competing policy recommendations and cause misalignment with the EPA. Therefore, this section not only presents the policy advice for Ireland emerging from the national reports, but also analyses instances in which the ERC advice does not match that of the OECD.

PISA 2000

To improve Reading outcomes, the ERC recommended evaluating existing policies to help students with serious difficulties. It also suggested that a new syllabus for the Foundation level of the JCE[18] should be developed and more students should be encouraged to take it (Shiel, Cosgrove, Sofroniou, & Kelly, 2001a, p. 18). For Mathematics, the ERC highlighted the poor performance of higher-achieving students, and it argued that even though a syllabus revision emphasising more 'realistic mathematics' was 'premature' (p. 15), future developments 'should take account of differences […with] the content and format of the PISA assessment of mathematical literacy' (p. 18). For Science, the ERC argued that a science module should be available in Junior Cycle to complement the soon-to-be implemented science primary curriculum. With a similar rhetoric to the OECD's, the ERC warned that the current good performance 'should not induce complacency' (p. 19).

Most of these recommendations are in line with the EPA; however, the idea of encouraging low-performing students to take Foundation rather than Ordinary level, while informed by PISA findings, did not challenge systemic streaming as advocated by the OECD. This idea was reiterated two years later:

> 'It is significant that 28% of students taking ordinary level syllabus […] achieved scores at or below Level 1. This […] points to a need to review the purpose and focus on foundation level English and the basis upon which students make choices about which level to take. Such a review would need to take into account the number who did not achieve grade D or higher in Ordinary level English […]

[18] The easiest of the three examinations (and related curricular pathway) available to students. The other two levels are Ordinary and Higher.

> coupled with the large proportion of Ordinary-level students achieving Level 1 or below on PISA reading literacy.' (Cosgrove et al., 2003, p. 37)

In both instances, the ERC advice was not built upon the OECD recommendation of reducing student segregation, but rather attempted to make existing streaming more efficient or less unfair. The motives behind this recommendation are not known, but it evidences a disconnect between the use of the "descriptive" aspect of PISA (achievement statistics) over its normative component at the national level, as argued in Chapter 1.

PISA 2003

Only a summary report was released by the ERC in 2004. The advice for Reading was similar to that in 2000 and focused on developing policies to help very low achievers. In Mathematics, PISA outcomes (which were average but had neither increased nor decreased) were instrumental for advocating a curricular revision. Changes in the mathematics syllabus were no longer 'premature' as the ERC pushed for greater alignment with the principles of Realistic Mathematics Education (a movement emerged in the Netherlands in the early 1970s, see Van den Heuvel-Panhuizen & Drijvers, 2014).

The ERC did not fully embrace the PISA curriculum. It recognised that the PISA Mathematics content had limitations. However, PISA results were taken as a good measure of realistic mathematics ability because 'the framework for PISA mathematics is grounded in the Realistic Mathematics Education movement' (Cosgrove, Shiel, Sofroniou, Zastrutzki, & Shortt, 2004, p. 40). Therefore, outcomes from PISA were a sign that mathematics teaching in Ireland did not place sufficient emphasis on real-life situations and problem-solving (Cosgrove, Shiel, Sofroniou, et al., 2004). The solution was for Irish mathematics curriculum and practice to move towards PISA in order to move towards Realistic Mathematics Education.

This is a use of PISA that was discussed in Chapters 1 and 2, as a means of problematising a situation and suggesting solutions, but it is also an interesting case of policy alignment. There is a feeling that the PISA curriculum was endorsed *only because* the ERC believed that it was a gateway for a different kind of mathematics instruction. Rather than aligning to the EPA, the ERC was aligning the EPA to its views. Consequently, it is possible that other findings from the international reports were noticed because they agreed with the ERC's priorities. These may include recommendations on classroom disciplinary climate and home educational processes, the continuation of existing interventions addressing socio-economic disadvantage, and the development of new ones to help single-parent families.

The ERC also noticed that the students not taking the science JCE performed worse in PISA and were predominantly female; therefore it advised, in line with the EPA, that 'all First-

year students, and especially females, [should be] well informed about the benefits of choosing science as a subject' (Cosgrove, Shiel, Sofroniou, et al., 2004, p. 43). The positive OECD finding (2004b, p. 121) whereby Ireland was the only country in which males did not express higher motivation or engagement in Mathematics, instead, was not commented on.

It is unclear why some PISA findings captured the attention of the ERC and others did not. As mentioned earlier, analysing the production of knowledge by national policy brokers was outside the scope of this research because if the EPA "works", it should work regardless of whose advice a country is aligning to. Nevertheless, it is a reminder that the OECD may have only tangential influence on national policymaking when the voice of national advisors is sufficiently strong.

The 2005 symposium

A symposium aimed at 'second-level teachers, representatives of the Department of Education and Science and its agencies, lecturers and researchers in education at third-level institutions' was held 'to reflect on the outcomes of PISA 2003 in Ireland, in terms of their implications for practice, research and policy development' (DES & ERC, 2005, p. 2). Proceedings from the symposium (ERC, 2005) are illustrative of what policy areas captured the stakeholders' interest at that time. Intervention strategies showed a remarkable alignment with the EPA.

Three problematic areas were identified in Reading: a sizeable percentage of very low achievers, socio-economic disadvantage and gender gap. To increase achievement, the introduction of early support mechanisms paired with teacher and whole school CPD was suggested, and so was participation in PIRLS as a way to monitor performance at the primary level. To tackle disadvantage, lessons from Nordic countries as well as school allocation policies were mentioned. Finally, participants discussed how to get boys more engaged with reading through teaching practices, specific assessment formats and more 'appealing' literary genres in the curriculum.

In Mathematics, delegates seemed to agree that a greater focus on problem-solving skills and real-life contexts was needed to raise student achievement. PISA was recognised as a valuable instrument. Some delegates asked whether the national curriculum should 'be altered to reflect something more of what PISA assesses'—even though it was unclear whether PISA tested 'literacy or ability' (ERC, 2005, p. 6). PISA managed to problematise mathematics education in Ireland. It was widely agreed that there was 'a need for a major debate on the kind of mathematics curriculum that [… stakeholders wanted] in Ireland' (*ibid.*).

Promotion, uptake and the relationship between the syllabus and the real world were mentioned also in respect of Science. It was argued that the recently-revised 2003 curriculum 'was more of a response to concerns about attitudes towards the hard sciences and uptake at

Leaving Certificate' rather than an implementation of an underlying 'coherent philosophy' building on the primary curriculum (ERC, 2005, p. 17), which would ensure a smoother transition between primary and secondary education. A lack of resources for science teaching was also lamented.

PISA 2006

With the third cycle, trend analyses became an important part of the national reports. It is interesting to note how lack of change in Irish results, which relegated the country to the background in the international report, was instead highly relevant at the national level, given the many curricular reforms that were taking place in those years (see the "Changing Curricula" section below).

Following a slight decrease in Reading 2003, a non-statistically significant improvement in 2006 was almost welcomed with relief by the ERC, as it was 'sufficient to ensure' that the overall 2000–2006 trend was not negative (Eivers, Shiel, & Cunningham, 2007, p. 35). The 2006 national report was also more critical of PISA compared to previous years. It was suggested that the 2000–2003 decrease might have been due to 'technical difficulties with a small number of items' (*ibid.*), and that the gender gap could be partly explained by the assessment contents and format. Policy recommendations were still highly aligned with the EPA but cited national evaluations (e.g., Inspectorate, 2006) rather than PISA reports. They included delaying ability grouping, increasing opportunities for collaborative teacher planning, emphasising formative assessment and raising teacher expectations of lower achievers.

In Mathematics, lack of change was used to critique the backwash effect of existing assessments on instruction and to promote the realistic approach endorsed by Project Maths, a curricular intervention that was being developed at the time.

In Science, lack of change was met with disappointment, given recent changes in the primary and secondary curricula that had made them more closely aligned with PISA. This also led to further criticism of PISA: 'Section 1B of the [revised Junior Certificate Science Syllabus …] is not represented by *any* PISA items. While it is impossible to provide in-depth coverage of all the main branches of science, it is ironic that an assessment of real-life science ignores the development of life itself' (Eivers et al., 2007, p. 35, emphasis in the text).

A few pages later, however, PISA was again the reference point to critique the effectiveness of the Ordinary level of the curriculum: it 'does not provide [… students] with any additional knowledge of science (as measured by PISA) and therefore raises the question of what scientific skills and knowledge the course does provide, and what PISA assesses' (p.

37). It looks as if some aspects of PISA were critiqued and some praised depending on whether they agreed with the ERC view of what was valuable education.

Most recommendations for science were integrated into the full report one year later (Eivers, Shiel, & Cunningham, 2008) and focused on monitoring curricular implementation, increasing technical support to science teachers, promoting science and science-related activities to parents and students, and motivating students. The critique to the Ordinary-level syllabus was complemented by the advice that the motivation and attitudes of students taking that pathway should also be examined.

The 2008 symposium

The proceedings from the 2008 PISA Symposium are not available, but it was possible to access the concluding remarks of the Assistant Chief Inspector (Murtagh, 2008). In line with the ERC, Murtagh praised Project Maths as the solution to Irish performance and uptake. The author supported his argument with a joke comparing the existing syllabus to an old dog that had to learn new tricks because 'the old tricks are not compatible with Windows Vista' (p. 6). Windows Vista was the newest Microsoft operating system at the time and the implication seemed that the mathematics syllabus had to adapt to keep up with the requirements of contemporary society or the demands of PISA.

PISA 2009

PISA 2009 was characterised by a large performance drop in Reading (−31 points) and Mathematics (−16 points), which turned Ireland into a negative example in the international report (OECD, 2010d). This captured all the attention and pushed other findings to the background: the decline in disciplinary classroom climate, as well as positive aspects such as the consistently low between-school variance, went uncommented. This gives support to the argument that the greatest achievement of the OECD with PISA has been to popularise the idea that scores and rankings matter (Grek, 2009; Martens & Niemann, 2010; Martens, Rusconi, & Leuze, 2007). Had this not been the case, the ERC would have granted other indicators the same relevance as it did to the decrease in scores. Instead, on 13 January 2011, representatives from the ERC, the DES and the Inspectorate were questioned by the Joint Committee on Education and Skills (2011) about the results: the title of the discussion was "Ireland's ranking on PISA", which is revealing of the policymakers' priorities.

The outcomes of PISA 2009 resulted in the production of five additional reports (Cosgrove & Moran, 2011; Cosgrove, Shiel, Archer, & Perkins, 2010; Cosgrove, 2011; Perkins, Cosgrove, Moran, & Shiel, 2012; Shiel, Moran, Cosgrove, & Perkins, 2010), plus two from independent sources (Cartwright, 2011; LaRoche, Cartwright, & Statistics Canada, 2010), as

well as two journal articles (Cosgrove & Cartwright, 2014; Eivers, 2010), all determined to explain (to stakeholders) the plausible causes for the falling scores.

It is telling that, in these reports, the ERC dedicated extensive effort on the validation of PISA. Some discontent had been expressed after PISA 2006, but this was the first time that the validity of PISA was so explicitly put under scrutiny[19]. The ERC identified three main factors behind the 2009 drop: student engagement, demographic and curricular changes. Student engagement referred to an increase in the percentage of missing and non-reached responses, especially in link items, which might have introduced DIF and biased the trends downwards (Cartwright, 2011; Cosgrove & Cartwright, 2014). This finding is further evidence of the effect of linking error on country outcomes critiqued in Chapter 4.

Demographic changes in the sample were represented by an increase in low-achieving students: first-generation immigrants, students with special education needs and students at risk of early school leaving (Cosgrove & Cartwright, 2014; Perkins, Moran, Cosgrove, & Shiel, 2010). Many of these findings were not discussed from a policy perspective. It is striking that a social issue (the declining performance of first-generation immigrants) and an educational success (the reduction in the proportion of school leavers) were instrumental for the analysis of PISA trends but were not treated as policy-relevant results in their own right.

Claims about the impact of the science curricula on the observed PISA outcomes are similarly puzzling. The ERC suggested that reforms of the primary and secondary curricula 'may have mitigated the effects of changes in demography and sampling that might otherwise have lowered performance in science in PISA 2009' (Perkins et al., 2010, p. 58). The ERC did not have comparable data from national assessments showing improvements in science performance. However, it did have: evidence of lack of impact of the primary and secondary curricula in English and Mathematics (e.g., Perkins et al., 2010); evidence of moderate implementation of the primary (Varley, Murphy, & Veale, 2008b) and Junior Cycle science curricula (Eivers, Shiel, & Cheevers, 2006); and evidence of the stability of Irish performance in science from all previous international assessments. In other words, it had more information suggesting that Science results should *not* have improved, rather than the opposite.

Perhaps the ERC had additional evidence that warranted its positive interpretation regarding the role of the science curricula (indeed, Science scores increased again in 2012). An alternative explanation is that the revised science curricula were much more aligned with

[19] From informal conversations with ERC staff, it emerged that scores had been stable for so many years in previous national and international assessments that the ERC thought a sudden drop of that magnitude was unlikely. This author noted that if outcomes were so predictable there would be little incentive for participating in PISA. The ERC staff counter-argued that the usefulness of PISA came from its rich databases rather than from mean country estimates, even though these latter drew the attention of policymakers.

PISA in terms of real-world contextualisation and problem-solving—an approach that the ERC endorsed (at least in mathematics). This may have led the centre to be slightly biased towards its effectiveness.

Analysis of selected policies

The analysis in this section covers reforms of three aspects of the Irish education system: curriculum, evaluation and assessment, and equity. Curricular and equity reforms have the potential to impact on PISA outcomes directly. Interventions on systemic evaluation and assessment were included as they represent the typical accountability policies that some research strands claim to be driven by globalising forces such as PISA. The case studies provided a good opportunity to understand whether these policies were actually the product of global forces, what role the local context had in shaping them, and how they could be linked to the EPA.

The analysis of each policy follows a prescribed format including context and influences, contact points with the EPA (or PISA, in case of curricular reforms), implementation, and (when possible) impact. The first step contributes to an understanding of how "international" discourses are translated on a national level. The second step helps to differentiate the level of goals and values (acknowledgment) from that of alignment (see Chapter 2): a policy may be influenced by international discourses or PISA outcomes in its general ideas or objectives and still not be aligned with the EPA in its specific modes of intervention. The third step links enactment of the EPA to its ability to influence student outcomes.

Evidence of alignment with the EPA and effectiveness from each policy feeds into the broader end-of-chapter conclusions which, in turn, contribute along with the cross-country evidence from Chapter 4 to answering the second and third research question.

Changing curricula

Mathematics: the revised Junior Certificate Mathematics Syllabus (rJCMS), 2000

Context and influences

After the introduction of the Junior Certificate in 1989, a mathematics committee was established and asked, in late 1994, 'to critique the Junior Certificate mathematics syllabuses […] with a view to introducing some amendments if required' (DES & NCCA, 2002, p. 6). A final draft was approved by the NCCA council in May 1998, and the revised syllabus (the rJCMS) was introduced in September 2000 for first examination in 2003.

'The drive to change came from the Inspectorate' (I4[20]), but many other stakeholders were unhappy with the state of the system. The exams were highly predictable, so much that newspapers published hints on how to answer specific questions before they came out, without any resistance from the examination board (I4, see also Lubienski, 2011). There was also an 'enormous cultural drive' (I4) to teaching to the test and this was partly due to teachers' inadequate subject knowledge—an issue that is still relevant today (I11). 'Junior Certificate mathematics teachers [… were] usually not subject specialists in that they [… taught] other subjects as well as mathematics' (Lyons, Lynch, Close, Sheerin, & Boland, 2003, p. 5; citing Oldham, 2001). As a result, they favoured abstract, procedural knowledge (Beaton, Mullis, et al., 1996).

Contact points with PISA

In its intentions, the rJCMS focused more than its predecessor on mathematical understanding and problem-solving in real-life contexts, in line with Realistic Mathematics Education, but its main source of inspiration was the recently-revised primary curriculum (NCCA, 1999). Merriman et al. (2014) argued that Realistic Mathematics Education underpins both the primary curriculum and PISA. This would provide a *trait d'union* between primary curriculum, secondary curriculum and PISA, but it is also where the similarities stop.

The primary curriculum borrowed from Dewey's (1915) idea of balanced education (Breacháin & O'Toole, 2013), and its guiding principle was 'celebrating the uniqueness of the child' (NCCA, 1999, p. 8). Of course, the primary curriculum had been developed in a period in which international LSAs were 'receiving attention' (DES, 1995, p. 203). But the educational goals it promoted lacked references to concepts such as "ability", "growth" or "econonomic development", and its international vision was limited to the role of Ireland within the European Union. It is also difficult to argue that the move towards realistic mathematics and problem-solving was to promote 21st century skills, given that the former refers to a movement initiated in the 1970s and the latter had been in the Irish agenda since the 1980s (I4).

In summary, the rJCMS had little alignment with the EPA/PISA, but acknowledged its importance by suggesting that the PISA model could form the basis for future reviews of the mathematics assessment practices (DES & NCCA, 2002).

Implementation and impact

The rJCMS was accompanied by targeted CPD focusing on realistic mathematics education, but the impression at the time was that the reform had little impact on classroom practice

[20] Interviews with Irish experts are coded "I" (for "Ireland") plus a sequential number. In some occasions, it was felt that even this system may compromise the anonymity of the participant, in which case the number was replaced with a bar ("I-").

(Lyons et al., 2003). As noted earlier, the ERC suggested further curricular revisions as early as after the first PISA cycle, one year after the introduction of the rJCMS. The rJCMS 'had to be done' (I2) but its effectiveness was severely hampered by the fact that the examinations (and therefore teaching) did not change. Stakeholders knew that to achieve results 'everybody needed to be engaged' and 'the whole problem' needed to be tackled (I2). This was the approach that became Project Maths (below in this chapter).

Science: the revised Junior Certificate Science Syllabus (rJCSS), 2003

Context and influences

Like the mathematics syllabus, the science syllabus (rJCSS) was prepared by a subject committee as a review to the 1989 curriculum (NCCA, 2003). The syllabus was devised in a period of substantial investment in science and technology by research councils and higher education institutions, but concern was expressed about science uptake and teacher training (MoE M. Martin, 1999a, 1999b; Task Force on the Physical Sciences, 2001). The rJCSS was introduced in the same year as all primary school pupils began to study the subject as part of the social, environmental and scientific education syllabus (DES & NCCA, 2006).

Contact points with PISA

The rJCSS was much more responsive to international trends and the knowledge economy framework than the rJCMS, but so was science education in Ireland in general.

O'Leary (personal communication of 19 November 2014) argued that the perceived poor performance in the IAEP2 of 1991 was 'a real wake-up call for science'. The science domain of the primary curriculum was 'largely based on the content and cognitive domains found in TIMSS 1995 and is considerably different to its predecessor, Curaclam na Bunscoile' (C. Murphy, 2013, p. 179; also I11). The Task Force on the Physical Sciences (2001) argued that 'Science should be included as a core component in the education of all students [also because of] its inclusion as one of the three literacy domains examined by the OECD as part of the Programme for International Student Assessment (PISA)' (p. 40). Following these recommendations, the National Policy and Advisory Board for Enterprise, Trade, Science, Technology and Innovation (Forfás) launched, on 30th October 2003, the Discover Science and Engineering Programme, a national awareness intervention that brought together a range of pre-existing campaigns (Forfás, 2004).

Curricular changes introduced with the rJCSS improved 'in many ways […] the alignment between the Junior Certificate Science Syllabus and PISA, in terms of moving closer to the concept of scientific literacy as defined by the PISA assessment framework. In fact, the definition of scientific literacy used for the PISA 2000 assessment […] was cited as part of the rationale for a syllabus revision' (Eivers et al., 2006, p. 4; see also NCCA, 2008). One

informant said: ‘Cynical people say it’s to prepare students for PISA, because they [PISA tests] are carried out one year later [compared to the science JCE]… but that’s cynical people’ (I10). Stakeholders’ perception of a strong curricular alignment between the rJCSS and PISA might explain why the ERC was disappointed when PISA 2006 “failed” to test one area of the rJCSS (Eivers et al., 2007).

Implementation

To facilitate the implementation of the rJCSS, a Junior Science Support Service was introduced with CPD opportunities involving ‘six one-day seminars over the first three years of the syllabus implementation’ (Eivers et al., 2006, p. 7). Typical attendance ‘comprised 75-80% of existing Junior Cycle teachers’ (Varley et al., 2008b, p. 37). However, Eivers et al. (2006) reported that ‘the revised syllabus has achieved some, but not all, of its aims’ (p. 29), because ‘while many teachers support the syllabus as theoretically constructed, there are difficulties with its implementation’ (p. 28). These difficulties involved lack of differentiation, an excessive use of textbooks, a teacher-centred approach and few opportunities for students to engage in practical work (Inspectorate, 2008, drawing from 63 inspections carried out between 2004 and 2006).

Following teacher feedback, the syllabus underwent minor changes (NCCA, 2008). After a few years, Varley, Murphy, & Veale (2011) found evidence of a greater emphasis on practical work in 15 schools. However, the authors also reported a lack of consistency between the primary and post-primary science experiences, ongoing traditional teaching, and ‘a neglect of certain areas recommended in the new syllabus’ (p. 295).

Teacher training was another issue. Using data from 150 pre-service teachers, Murphy & Smith (2012) reported that student teachers had not acquired ‘sufficient understanding of scientific background knowledge to implement the science curriculum in a sufficiently competent manner’ (p. 89).

Impact

As in 2009, Perkins et al. (2013) argued that the introduction of scientific education in the 1999 primary curriculum and the rJCSS ‘may have contributed to the significant increase in science achievement observed in Ireland in [PISA] 2012’ (p. 161). Yet the evidence above provides only limited support to the hypothesis that the syllabus was implemented consistently. Even assuming that it did and that this would have been sufficient to affect PISA outcomes, it is unclear why the performance increase was confined to Second- and Third-Year students and students in the 10th percentile, whereas Transition and Fifth Year students did worse.

Recent developments: The Literacy and Numeracy Strategy and Project Maths

Some interventions have taken place too recently to be thoroughly evaluated, but they are illustrative of recent trends in curricular development.

The National Literacy and Numeracy Strategy (NLNS), 2011

In July 2011, a strategy called "Literacy and Numeracy for Learning and Life: The National Strategy to Improve Literacy and Numeracy among children and Young People, 2011–2020" was launched in Ireland. The name is reminiscent of OECD's *Knowledge and skills for life,* and indeed some think that the strategy was 'completely informed by PISA 2009' (I6). More moderate statements argue that it 'was informed by evidence from [… PISA 2009] and findings from Inspectorate reports and other sources that suggested that the teaching of literacy and numeracy needed to be strengthened' (Inspectorate, 2013, p. 15). In any case, the NLNS provides an example of enactment of the EPA in Ireland.

The NLNS draws heavily from OECD recommendations and improvement in PISA is one of the targets set for secondary education (DES, 2011b, p. 18)—a choice considered by some 'embarrassing… almost' (I9), since Irish governments have a tendency to miss these kind of objectives. The importance of literacy and numeracy had already been recognised in the primary curriculum (Breacháin & O'Toole, 2013; O'Donoghue, 2002), but they were part of a broader set of academic and socio-emotional skills to be taught. With the NLNS, the holistic approach of the primary curriculum disappeared and cuts were recommended to 'desirable but ultimately less important activities' (Breacháin & O'Toole, 2013; DES, 2011, p. 15). What had changed?

On the one hand, a narrative of crisis (Apple, 1988; Takayama & Apple, 2008) had entered the education policy arena, riding on the back of the actual economic crisis of 2008. As a result, the narrow focus of the NLNS was presented as a 'difficult choice[…]' that 'ha[d] to be made' because of the 'extremely difficult' economic circumstances, in order to 'ensure the most efficient use possible of available resources' (DES, 2011b, p. 15). The language of crises and opportunities is a staple of PISA reports, therefore outcomes from the assessment could easily resonate in Ireland. The new government, which came into power in 2008, needed 'to be seen to respond to PISA results' (Breacháin & O'Toole, 2013, p. 413), and this response took the shape of the NLNS.

On the other hand, the system was ready for a change (see Steiner-Khamsi, 2006, on timing). Calls for greater focus on literacy and numeracy had started 'well before PISA 2009' (I9). Kellaghan et al. (1995) had recommended 'paying particular, though not exclusive, attention to literacy and numeracy skills' (p. 66). Archer & Shortt (2003) had critiqued 'the absence of a clear prioritisation of literacy and numeracy' in schemes for tackling disadvantage (p. 19). When the PISA 2009 results came out, the DES 'was in a stronger

position to argue their [its] case' (I11): 'when you are trying to push a policy, it does not hurt that your country is slipping in the PISA' (I7). Teacher unions were more ready 'to buy in the policy than if PISA hadn't happened' (I9).

In spite of its extensive instrumental use by the government and the unions, one informant thought the way PISA interprets mathematics as literacy has educational value: 'students need to be able to mingle their literacy with their numeracy and need to know when to use them' (I1). Instead, Irish students treated 'the wordy part' as disconnected from the procedural part, 'as if they were being asked opinions, they couldn't see that it was still maths' (I1).

Project Maths, 2012

Context and influences

The three principles underpinning Project Maths are Realistic Mathematics Education, knowledge economy and problem-solving: 'It involves empowering students to develop essential problem-solving skills for higher education and the workplace by engaging teenagers with mathematics set in interesting and real-world contexts' (NCCA, 2015). '*Project Maths* is a work in progress' (NCCA, 2012, p. 1) that 'sets out to change the traditional didactic approach' (McMorris, 2008): 'we are changing a culture [… ;] we are moving, as many European countries are, towards more emphasis on applications and producing mathematically literate people rather than people who are able to reproduce information' (McMorris, 2008).

Project Maths was more than a curricular reform, it was meant to be a paradigm shift in mathematics instruction and assessment. It was a cultural change that had been brewing for two decades. Until the early 1990s, the DES claimed the education system was excellent and the attitude towards assessment was: 'let's not measure it, in case it isn't' (I4). It was a tried and true paradigm, and in this lay its strength and weakness (Oldham, 2001). Following a generational change, the DES became more open to data-gathering approaches like inspections and comparative surveys (I4). TIMSS 1995 critiqued a traditional approach to teaching and assessment and teacher over-reliance on textbooks. These findings informed the first curricular reform (the rJCMS) described in the previous section: 'perhaps TIMSS 1995 helped to ask questions' (I4).

As seen earlier, after the first and the second PISA cycle, the ERC attacked current mathematics education and the rJCMS. Critiques were not limited to national PISA reports. In 2003, the influential book *Inside Classrooms* (Lyons et al., 2003) corroborated suspicions that teaching practice was exceedingly similar in schools across the country (I1). Interestingly, Lyons et al. (2003) were much better-disposed towards PISA than towards ETS and IEA studies. The book opens with a reminder about the importance of mathematics for the knowledge economy, and the PISA objectives are readily endorsed at face-value: all

international studies suffer from limitations, 'not least because they assess only limited aspects of the curriculum', but PISA is an 'exception', because it aims 'at assessing how well students can use mathematics to solve realistic problems' (pp. 9–10).

From its end, the ERC continued to attack teachers' over-reliance on traditional text-book problems—unlike PISA, 'which places a strong emphasis on solving mathematics problems set in real-life contexts' (Shiel, Surgenor, Close, & Millar, 2006, p. 154, following the 2004 national assessment results). Teachers were said to be lacking 'confidence in using constructivist approaches in teaching problem solving' (*ibid.*). The ERC was particularly critical of the role of existing examinations as an obstacle to changing instructional practices (Cosgrove, Shiel, Oldham, & Sofroniou, 2004).

The NCCA added to the pressure. In *International Trends in Post-Primary Mathematics Education*, Conway & Sloane (2005) extensively argued that the existing system was inadequate to meet global challenges. A major reform, of assessment practices in particular, could 'produce a significant ripple effect right through secondary education' (p. 248). A consultation on mathematics education was started. First, within the NCCA, which comprises '25 members […] appointed by the Minister and [coming] from organisations representing teachers, school managers, parents, employers, trade unions and interests in education' (NCCA, 2009, p. 6). Then, between November 2005 and January 2006, it was also open to the public.

The first discussion paper from the consultation (NCCA, 2005) adopted not only the OECD definition of literacy, but also the rhetorical style of its reports. From the second discussion paper, it emerged that many agreed 'that more concrete experiences of mathematics in real-world contexts [were] needed for primary and junior cycle students' (NCCA, 2006, p. 15). The NCCA consultation papers were well-received by the Royal Irish Academy Committee for Mathematical Sciences (2006), the Irish Mathematical Society (O'Reilly, 2006) and the Irish Maths Teachers Association—'the only Maths organisation that is permitted to have an input into the development of the programme' (2013, p. 3).

Isolated criticism came from academics at University College Cork who were 'very concerned about the influence of the PISA philosophy of mathematical education which seems to be heavily influencing Project Maths' (Grannell, Barry, Cronin, Holland, & Hurley, 2011, pp. 3–4). Others argued that 'teachers were mostly on the outside of the consultation process because they do not want to be consulted: they want to maximise points [at JCE]' (I3).

After PISA 2006, the ERC related the visions of PISA and Project Maths and suggested they were not opposite to currently-taught content: 'Ultimately, it may be a case of adjusting the current Junior Certificate syllabus and examinations to address some of the apparent

shortcomings identified by PISA and other studies, while at the same time retaining the most important content' (Shiel, Perkins, Close, & Oldham, 2007, p. 47).

The first document identifying the new initiative as "Project Maths" was produced in May 2007, and in July the NCCA submitted a report to the Department of Education (I-). In January/February 2008 the NCCA council asked for a commitment to the Department of Education of multi-year funding of intervention, which was agreed in March/April of that year (I-).

Given that discontent had been mounting for some time, experts might be correct in saying that 'there was not a *huge* impact of PISA, only *some* impact' (I2); that PISA was 'another spike in the wheel that maths teaching needed to change' (I1). Other motivations for Project Maths came from performance at Leaving Certificate, dropout rates, and poor mathematical preparation of students when they entered third-level education (I4). However, it is clear that the issue at the time was not whether, but to what extent mathematics education in Ireland should 'move toward a PISA-like approach' (Conway & Sloane, 2005, p. 213)—or 'buy into PISA', as participants to the PISA Symposium put it (ERC, 2005).

Contact points with the EPA

On paper, Ireland "bought into" PISA—or, more precisely, into the EPA—to a considerable extent. Apart from the overall vision of mathematics education, Project Maths was meant to increase formative assessment, CPD, teacher collaborative action research and parental involvement in the teaching of mathematics (NCCA, 2006). Curricular alignment with PISA was already high before the reform, but mismatching areas were not necessarily amended; for instance, the new syllabus paid little attention to concepts of space and shape, although they are an area in which Ireland has consistently underperformed (I10). The boldest move towards PISA, and the greatest rupture with the past, however, was in the assessment format, which was very much PISA-like. This made Irish students more familiar with this kind of test (I2; I4; Merriman et al., 2014).

Implementation

The reform was firstly piloted in 24 schools in 2008 for testing in 2010 (Leaving Certificate) and 2011 (Junior Certificate; McMorris, 2008; NCCA, 2011). National rollout began in September 2012 for Junior Certificate examination in 2015, whereas the Leaving Certificate underwent further revisions and was implemented nationwide one year later (NCCA, 2014, but see also Jeffes et al., 2012, p. 18, for a summarising table).

Preliminary reports found mixed-to-positive results in terms of programme implementation (Jeffes et al., 2012, 2013; Lubienski, 2011; Merriman et al., 2014). There was overall support for the reform which still had to translate into classroom practice and student work. Mathematics textbooks in 2010/2011 were 'a genuine attempt to match the intentions

of Project Maths' but they fell short of the standard needed to support it (O'Keeffe & O'Donoghue, 2011, p. v).

Some doubts were raised about the ability of the intervention to keep the momentum in light of some resistance towards the new examination format. Almost immediately, there was a 'national outrage over a mock paper' which was 'out of proportion' but was nevertheless sufficient to make some policy makers wonder whether it would be worth 'rolling-back' to a better-known and 'safer' format (I1).

Impact

Impact evaluations were also more positive than negative. Jeffes et al. (2013) reported 'emerging evidence of positive impacts on students' experiences of, and attitudes towards, mathematics' (p. 6) but no overall change in achievement using a purposefully-developed test comprising TIMSS and PISA items. Merriman et al. (2014), in contrast, compared the performance in PISA 2012 between the 24 schools participating to the initial phase of Project Maths and non-participating schools. They found that, once socio-economic and demographic variables were accounted for, attending a school participating to Project Maths was associated with 'a 10-point advantage on print mathematics over students in Non-initial schools' (p. xi), and that stronger improvements could be seen particularly in PISA areas in which Irish performance had been weaker in the past. Still, considering that the PISA standard deviation is 100 points, a 10-point difference represents a small effect size (0.1).

Changing evaluation and assessment

Assessment in primary education

In primary education, pupils undergo standardised assessments of progression. Moreover, every five years, a cohort sits an ERC-developed low-stakes national assessment of reading and mathematics (NAMER). The NAMER was commissioned 'to establish links […with] the Third International Mathematics and Science Study (TIMSS)' (Shiel & Kelly, 2001, p. v) and to this date it borrows from TIMSS and PIRLS. This section focuses primarily on the standardised tests, since they bear higher stakes for teachers and pupils.

Context and influences

In the early 1990s, during the national debate culminating in the Education Act, 1998, 'proposals relating to the assessment of pupils by their teachers at the end of first and fifth class' started to surface in the Green and White paper (INTO, 2008a, p. 14), as well as in NCCA publications (1993). However, many stakeholders thought that the proposals 'would cause considerable disquiet' because such assessments 'could devalue the teaching process'

(INTO, 1997, p. 5); therefore, mentions of standardised testing disappeared from the policy discourse for some time.

In 2000, the *Drumcondra English Profiles*—'cumulative records of achievement […] that are based on teacher judgements' (Shiel & Murphy, 2000, p. 1)—were introduced. Teacher unions claimed the Drumcondra profiles 'caused grave concern among teachers' (INTO Education Committee, 2004, p. 55), yet 'nobody pushed them' (I10) and to this date they have never been compulsory.

In 2004, standardised testing started to be associated with accountability practices in governmental discourses; there was 'a need to have aggregated assessment data for decision-making, identifying progress and the allocation of resources' (INTO, 2008a, p. 17). Circular 0138/2006 (available from www.education.ie) made standardised testing of reading and mathematics between the 1st and 2nd Class and the 4th and 5th Class compulsory from 2007. In the same year, national assessment guidelines were produced (NCCA, 2007).

This change of pace did not affect standardised testing only. The NCCA guidelines also included recommendations from an expert group on assessment for learning (AfL). The AfL framework was devised between 2005 and 2007 and had been informed by the works of Black and Wiliam (1998), the Assessment Reform Group (1999), as well as by 'the work of others in the area, such as Airasian, Stiggins, Hattie, Crooks etc.' (O'Leary, personal communication of 19 November 2014).

Standardised testing remained compulsory at two stages of primary education until 2011, when Circular 0056/2011 raised the frequency to 2nd, 4th and 6th classes from 2012 and required schools to report data to the Department of Education (see also Circular 0018/2012).

Contact points with the EPA

The EPA supports standardised assessments for diagnostic purposes as an early-warning mechanism to help identify and support struggling students (the AfL component). It also welcomes standardised testing for school accountability by public posting of achievement data, provided it is accompanied by greater school autonomy. However, it is argued below that Irish tests to date have served neither purpose.

Implementation and impact

When officially "implemented", standardised testing was nothing new for the great majority of teachers. Informally, 'assess[ing] the progress of pupils at the end of each year and particularly at the end of the final year at school' was 'routine practice' even before the Education Act (MoE M. Martin, 1997). Circular 0138/2006 estimated that '95% of schools already use such tests', whereas some informants mentioned values around 80%. Even if the government's rationale for the legislation was influenced by new public management ideas, the Circular did little more than formalise 'existing assessment practice' (INTO, 2008a, p. 37).

Eivers et al. (2010) reported that standardised tests were not widely used to establish school-level learning targets, to identify the pupils' strengths and weaknesses, to plan for differentiation at the classroom level or to inform parents, who are usually given verbal feedback or no feedback at all. There is no evidence of increase in testing (I9) and, given that the tests are not secured (I10), they cannot be used to track performance over time as they are affected by score inflation. CPD was also lacking, with only two days provided in 2007 that did not even cover the new guidelines, which were disseminated one year later (INTO, 2008b). Overall, standardised assessment 'appears to operate as a stand-alone activity […], and assessment data seems to be regarded more as a record of pupils' attainment rather than as a means of informing ways to improve and support their learning' (DES, 2012, p. 59).

With respect to AfL, schools are more aware that '[they] should be doing it', but the implementation of AfL strategies can still be categorised as 'emerging / sporadic' (O'Leary, personal communication, but see also Lysaght & O'Leary, 2013). Likewise, school inspections found improvements in the assessment of English between 2010 and 2012, but also 'significant shortcomings […] in almost one quarter […] of the English lessons observed' (Inspectorate, 2013, p. 45). Assessment 'is not satisfactory in 28% of mathematics lessons' and shortcomings 'tend to be in the area of formative assessment' (p. 48).

School inspections

The Inspectorate (a division of the DES) utilises many forms of inspection. Some are specific to secondary schools, but most are carried out at both the primary and the secondary levels. The most comprehensive are whole school evaluations (WSEs), assessing the quality of teaching and learning, leadership and school management, planning and self-improvement. But there are also thematic, follow-through, incidental (unannounced) and probationary teachers' inspections (Inspectorate, 2015).

Context and influences

The Inspectorate was not always as active, particularly at the secondary level. Between the 1970s and early 1990s it was burdened with administrative tasks and not allowed to interfere in school affairs by teacher unions (McNamara, personal communication of 20 November 2014; see also McNamara & O'Hara, 2012). This was observed also by the OECD (1991). At the time, the MoE defended the work of the inspectors—they were 'not "snoopers" but enablers' (p. 122)—but she conceded that their role could be reviewed.

Things began to change in the mid-late 1990s. On the one hand, the White Paper of 1995 separated evaluative and advisory functions in education matters: the Inspectorate was put in charge of the former and support services of the latter (INTO Education Committee, 2009).

On the other hand, there were demands for greater school accountability from parents and the business community, in line with European trends (DES, 2012).

The WSE strategy '[had] been in gestation since 1996' (Stack, 2005, p. 6) and was introduced alongside whole-school planning in 1999 with a pilot project in 35 schools. Since the external monitoring component of WSE encountered strong teacher resistance, the evaluation strategy had to be framed and presented as school self-evaluation (SSE, McNamara & O'Hara, 2005).

When the guidelines on WSE/SSE appeared (Inspectorate, 2003), external inspections had been relegated to the background but had not disappeared completely. This was a political victory for the Inspectorate and Chief Inspector, Eamon Stack, who was in charge between 1997 and 2009 and the person behind many of the interventions that increased the relevance of the Inspectorate during that period (National Council for Special Education, 2013). The evaluation framework had retained 'some of the key elements of external inspection central to the original proposal and resisted attempts to make the process one made up entirely of self-review and self-evaluation' (McNamara & O'Hara, 2005, p. 271). Not only this, but the project 'was endorsed by principals, teachers and inspectors and there was a general perception that the process was workable' (*ibid*.).

Contact points with the EPA

A presentation by the Assistant Chief Inspector (Egan, 2010) mentioned two influential sources for the Inspectorate's new evaluation approach: 'International reviews of education: e.g. PISA' and recommendations of 'international educational bodies' (p. 16) such as the OECD and the CERI. If one assesses the available evidence, however, it seems unlikely that the strategy could be specifically informed by PISA or by any *education* department within the OECD.

On the one hand, the only form of evaluation explicitly endorsed by the EPA is public posting of achievement data, which is illegal in Ireland (though yearly "league tables" are produced by some newspapers using other data, such as progression rates to university by school). On the other hand, the only source referenced by Egan (2010) was a peer review of public management policies (OECD, 2008a) not carried out under the supervision of the Education Secretariat or the CERI.

Of course, the OECD is broadly speaking not opposed to school accountability, capacity building and self-regulation, and apparently (I-) Eamon Stack thought the strong support of the OECD for school accountability was a very useful lever in his campaign to re-establish effective inspection. This, however, does not mean that WSEs in Ireland were based on evidence from PISA or any specific educational recommendation.

More likely, they were a 'refraction' (Sugrue, 2006) of international recommendations and European convergence on the same themes over the previous 20 years (McNamara, O'Hara, Lisi, & Davidsdottir, 2011) —a policy whose ambitions were international but whose actual design and development were idiosyncratic. In other words, the reform of the inspectorate seemed more an acknowledgement of the OECD goals than alignment with the EPA.

Implementation

The major limitation of WSE when first introduced was the fact that 'any kind of sensible and useful judgements [… required] data that in the […] system simply [did] not exist' (McNamara & O'Hara, 2005, p. 278). There was 'no way in which schools [could] hope to obtain any significant data on current performance and ways of improving' (p. 279). Moreover, schools did not have enough control over resources or staffing to address their shortcomings. Thus, the first attempts to implement WSE were described as 'data free evaluation' (McNamara, 2006, p. 581): highly supported by school staff but of dubious utility.

In 2010, a new policy document (Inspectorate, 2010) attempted to address these shortcomings. It placed a greater emphasis on gathering effectiveness data and acting upon it, and 'new procedures were put in place to enable […] schools to formally discipline and, if necessary, to dismiss under-performing teachers' (DES, 2012, p. 14). Inspections could now be triggered by school performance, though criteria and consequences were not specified (McNamara & O'Hara, 2012).

In 2012, WSE/SSE became a formal requirement to complement the implementation of the National Literacy and Numeracy Strategy (Circulars 0039 and 0040/2012), but once again an absence of a clear follow-up strategy could be noted. As of today, 'school inspection in Ireland remains largely subjective in nature. […] Objective measurements […] are not used in any systematic way to reach judgements, nor is there any system of sanctions or rewards for schools which fail to meet particular criteria' (Ehren et al., 2014, p. 57). Other research showed that school evaluation is felt in many cases as a pointless bureaucratic exercise that neither empowers schools, nor does it help them improve (McNamara et al., 2011).

The way policy affects teaching and learning in practice changes over time. In a personal communication (20 January 2016), McNamara noted that he is much more positive about inspection now than the quotes of 2011 might imply. This is because the most recent changes in inspection procedure in Ireland, he argued, have all combined to make inspection much more effective as both an accountability and formative tool. There is now a wider use of examination data; there are self-evaluations involving surveys of pupil and parent opinions; and WSEs were replaced by more risk-based and shorter and sharper inspections, requiring schools to formally respond in the form of action plans.

This is a reminder that policy implementation and integration is a slow but continuous progress, and that stakeholders' reactions and practices might change even substantively once certain tipping points are reached.

Changing equity

Delivering Equality of opportunity In Schools (DEIS), 2005

'Delivering Equality of Opportunity in Schools (DEIS) the Action Plan for Educational Inclusion, was launched in May 2005 and remains the Department of Education and Skills policy instrument to address educational disadvantage. The action plan focuses on addressing and prioritising the educational needs of children and young people from disadvantaged communities, from pre-school through second-level education (3 to 18 years).' (DES, 2014)

Context and influences

DEIS is the result of policies addressing disadvantage dating back to 1969, when the Rutland Street project attempted to affect the pre-primary and primary experiences of poor children (Holland, 1979; Kellaghan, 1977). Another intervention was the Disadvantaged Areas Scheme (DAS) of 1984, providing additional funding to primary schools designated as "disadvantaged" 'on the basis of the number of students from families in poverty as assessed by socioeconomic indicators such as unemployment, and possession of a medical card' (Lewis & Archer, 2002, p. 2).

The Rutland Street project evolved into Early Start in 1994 (see Lewis, Shortt, & Archer, 2011, for a recent evaluation). The DAS was the basis for other interventions, including the Home/School/Community Liaison scheme (HSCL) of 1993 (National Educational Welfare Board, 2013; Ryan, 1994), the Breaking the Cycle initiative of 1996–2001 (Weir, 2003) and its successor, Giving Children an Even Break (GCEB, Weir, 2004).

Each of these policies had its own history but eventually—following recommendations from the Educational Disadvantage Committee (2003a, 2003b) and ERC reviews of national policies and international literature (Archer & Weir, 2004; Weir & Archer, 2004)—they all flowed into the School Support Programme of DEIS. DEIS brought together, streamlined and improved existing measures to tackle disadvantage, including DAS, HSCL and GCEB; the Support Teacher Project; aspects of the Early Literacy Initiative such as Reading Recovery, the Junior Certificate School Programme Literacy Strategy and the Demonstration Library Project; and the School Completion Programme (DES, 2005).

DEIS is the result of national experiences that were sometimes affected by European standards and OECD recommendations. DEIS was informed by 'the Lisbon Strategy, the Agreed Programme for Government, the social partnership agreement Sustaining Progress,

the National Anti-Poverty Strategy and the National Action Plans against Poverty and Social Exclusion' (DES, 2005, p. 7). Some considered the focus of DEIS on literacy, numeracy and evaluation 'a very new idea for Ireland' (I7), and the ERC used the OECD definition of literacy in the reports informing the government's action plan for DEIS (Eivers, Shiel, & Shortt, 2004, 2005). Outcomes from an OECD review on early childhood care and education (OECD Directorate for Education, 2004) were used to 'support the development of future policy' in that area (DES, 2005, p. 34). Praise by the OECD of the effectiveness the HSCL scheme in fostering parental involvement (OECD, 1997c) was also mentioned, perhaps as a motivating example that good work could gain international recognition.

The greatest point of agreement between Ireland and the OECD, however, was in the very definition of "educational disadvantage". O'Sullivan (2005) argued that a 'mercantile' understanding of educational disadvantage had driven policy development at least since the mid-1990s (e.g., Kellaghan et al., 1995). Early school leavers and low achievers were 'positioned as deviant participants in a mercantile world' (O'Sullivan, 2005, p. 188). The Education Act (1998) brought the focus on the inequality of opportunities caused by differing socio-economic conditions and preventing students 'from deriving appropriate benefit from education' (§ 32(9)), but it was an isolated case. For instance, the National Development Plan 2000–2006 (Government of Ireland, 2000) suggested that marginalised communities should be encouraged 'to help themselves by identifying their own problems and working towards their solution in a planned and integrated way with the agencies of the State' (p. 196).

The framework endorsed by the Educational Disadvantage Committee for DEIS was devised and presented in November 2002 by 'Professor Peter Evans of the OECD' who 'complimented the government on taking the innovative action of setting up the [Educational Disadvantage] Forum, saying that it would [have been] difficult to convene such a meeting in other countries' (Educational Disadvantage Committee, 2003b, p. 9). The new framework was to be 'customer-driven', 'family-oriented', outcomes-based, preventative and intervening, 'decentralised' and funded by both public and private sources (Educational Disadvantage Committee, 2005, p. 11).

Contact points with the EPA

Because of its comprehensive nature, DEIS covers many areas of the EPA, including instruction time, the desegregation of migrant students, curricular changes, cooperation among schools and with parents, as well as the allocation of resources to schools and students (DES, 2005). There are also contact points with the specific EPA for Ireland. Disadvantaged schools were identified through a six-variable model measuring the percentage of: unemployment, single-parent families, travellers, large families, free book grants, and local authority housing. Of these, the performance of students coming from single-parent families was often

highlighted as an issue in OECD reports. This is not to suggest that it caused the variable to be included in the model (it had been part of the criteria to identify disadvantage since 1996, see Comptroller and Auditor General, 2006), but it means that both the Irish strategy and PISA were looking at the same population.

Implementation

Almost all reports on DEIS agreed that implementation levels were high (Inspectorate, 2011a, 2011b; Weir, Archer, O'Flaherty, & Gilleece, 2011; Weir & McAvinue, 2012; Weir, 2014). Weir & Errity (2014) noted that 'measures under DEIS exceed what was available under previous schemes and better reflect what has been identified as important in addressing disadvantage' (p. 15). An isolated critique came from Smyth & McCoy (2009), who argued that more than half of young people from at-risk families attended non-DEIS schools. The introduction of DEIS also meant that some schools received fewer funding than before, but principals were generally satisfied by funding levels (Smyth & McCoy, 2009).

Impact

Using baseline data collected in 2005 and 2007, Weir, Archer & Millar (2009) found that pupils from a poor background in rural schools performed better than those in urban schools, and that both urban and rural disadvantage students performed under the national average for reading but not for mathematics. The relative advantage of pupils living in the country was later attributed to a higher level of educational activities within the home and a lower level of unstructured free-time activities (Weir & McAvinue, 2013).

A summary of findings from three reports on the achievement of pupils in DEIS schools between 2007 and 2010 reported 'clear evidence that the DEIS programme is having a positive effect on tackling educational disadvantage' (DES, 2011a, p. 1; also Weir & McAvinue, 2013). Improvements were found in literacy and numeracy, attendance and home-school relations. In urban primary schools gains were 'particularly evident in the junior grades and in schools that have high levels of disadvantage' (Weir & Denner, 2013, p. 20). Even though the gap between DEIS and non-DEIS schools was still evident, DEIS schools improved at Junior Certificate level at a faster rate in the period 2008–2011 (McAvinue, 2014). Currently, there is a feeling that DEIS 'works better for pupils until the age of 12, less well for the 12–15 bracket' (I7, also MoE Quinn, 2014, p. 9).

Overall, the strengths of DEIS were identified in its integrated approach (Weir & Errity, 2014), the fact that it involves the school as a whole, and in its focus on literacy and numeracy in Junior Classes —an 'enormous emphasis' (Errity, 2014, p. 16) compared to previous interventions. Some also think it involves less paperwork for principals and school managers, but not everyone agrees on this point (I8).

Policy alignment of Irish policies to the EPA

The selection of policies above is just a sample of all the interventions that were analysed before and during the study visit. This section provides a comprehensive view of Irish reforms and the extent to which they were in line with the EPA. Information from this and the previous section then feeds into the end-of-chapter conclusions below.

The policy evaluation literature has a range of instruments to measure policy alignment between and within countries. Delta convergence (Heichel et al., 2008) is one of them. It attempts to evaluate changes in policy alignment between a country and 'an exemplary model, for example a model promoted by an international organisation or a frontrunner country' (p. 83). Delta convergence can be calculated when the policy options provided by the exemplary model can be quantified. Examples include taxes based on energy consumption or international regulations on the maximum amount of lead in petrol (Arts, Liefferink, Kamstra, & Ooijevaar, 2008).

Unfortunately, the exemplary model provided by the EPA does not set quantitative rules for policymakers. Following Arts et al. (2008) and Chapter 4, one could try to measure policy convergence by tracking the number of interventions broadly in line with OECD recommendations over time. The disadvantage of this approach is that it would not exploit the richness of the information gathered during the case study visits. It would not distinguish between acknowledgement, alignment or enactment, and it would not consider policy implementation.

Table 5.2 retains the concept of delta convergence while accounting for different degrees of policy alignment and implementation.

Table 5.2: policy alignment towards the EPA in Ireland

Year	T1	T2	E1	E2	E3	E4	E5	Q1	Q2	Q3	Q4	Q5	P1	P2	P3	Policies
1998						1				2						Education Act
1999															3	revised primary curriculum
2000													4			Education (Welfare) Act, rJCMS, Learning support, Equality Act
2001																Giving Children an Even Break
2002																Teaching Council Act, School Completion Programme
2003																rJCSS, State Examination Commission, Reading Recovery
2004			5													Equal Status Act, Education for Persons with SEN
2005											6					WSE/SSE, AfL, Disability Act, DEIS
2006																Síolta Quality Framework, National Behaviour Support Service
2007																Standardised testing
2008																Curriculum framework for Guidance
2009																
2010																Free pre-school year
2011																Literacy and Numeracy Strategy
2012																Project Maths

Colour coding:

Contrary	No policy	Acknowledgement	Alignment / Enactment	Implementation

The first column displays the year from the most recent overarching education framework before PISA, the Education Act (1998), until the last PISA cycle at the time of this writing. The T1–P3 columns are the EPA areas[21], and the last column lists the Irish policies under consideration by year of coming into force, when known, or publication.

The cell colours represent policy alignment: red is for policies that are contrary to OECD recommendations; white means that there is no national policy clearly attempting to intervene on that policy area; yellow stands for acknowledgement (interventions purporting or aspiring to be in line with the OECD goals); blue is for policies actually aligned with the EPA, whether intentionally or not and regardless of whether they managed to be implemented effectively; and green represents policies that are not only in line with OECD recommendations on paper, but for which there is plausible evidence of implementation.

Colour shades signal when policies first addressed the EPA (dark shade) and how long for (light shade). Once again, this does not take into account whether implementation was effective but only the fact that, at a certain point in time, a policy agreeing (dark blue/green) or disagreeing (dark red) with the EPA came into force, and that policy was "active" for a number of years afterwards (light blue, green or red). Notice that yellow comes in one shade only because it just serves to flag up purported but not actual alignment.

Finally, the numbered cells index some policies that are followed up in the following section to make more general comments.

[21] T1 = Increase overall time spent in education; T2 = Increase instruction time; E1 = Reduce disabled or immigrant students' segregation; E2 = Reduce the gender gap; E3 = Reduce the streaming / grouping / stratification of students; E4 = Reduce school competition; E5 = Help socio-economically deprived students; Q1 = Strengthen the teaching profession; Q2 = Allocate resources more efficiently or in a more equitable way; Q3 = Pair more autonomy with accountability measures; Q4 = Promote cooperation between parents, teachers, principals and schools; P1 = Help low-performing students; P2 = Standardise the educational outcomes; P3 = Review curricular priorities.

Trends in policy convergence towards OECD recommendations

Most cells in Table 5.2 are blue, suggesting that Ireland and the OECD generally agree on educational matters. The only contentious aspect is parental choice (point 1 in the table). The Education Act (1998) promotes 'the right of parents to send their children to a school of the parents' choice' (§ 6(e)), and although there have been interventions to increase equality of access, the segregating effect of high-SES parents sending their children to the "best" schools is still present. Because of the shrinking pool of secondary school applicants up until the early 2000s, school competition and polarisation even increased (Byrne & Smyth, 2010).

The Education Act also laid the bases for a proliferation of accountability policies (point 2), including WSE/SSE in 2005, public posting of school reports (but not achievement data) in 2006 and compulsory standardised testing in 2007. Three issues can be identified in this respect. Firstly, not all interventions drew from the EPA. Secondly, and relatedly, more accountability was not accompanied by increased school autonomy at least until 2010. Finally, there is only partial evidence that the "spirit" of the EPA was preserved when implementing these policies—i.e., that they were more than a bureaucratic exercise.

Points 3 and 4 serve as a reminder that some subject domains were more aligned with PISA after curricular revisions and more affected by measures to help low-performing students. An example was the scientific area of the revised primary curriculum (3, and see INTO, 2008b; Varley, Murphy, & Veale, 2008a). Likewise, learning support (4) was strongly biased towards English and relegated Mathematics—traditionally perceived as a less-prestigious subject—to a secondary role (Inspectorate, 2005; McCarthy & Burns, 2005; Surgenor & Shiel, 2008; Travers, 2010).

Similar observations can be made about equity, and in particular about equality of access (5). Recent legislation and interventions have been highly attentive to racial discrimination (Smyth, Darmody, McGinnity, & Byrne, 2009; Taguma, Kim, Wurzburg, & Kelly, 2009) and considerate towards the needs of students with learning difficulties (Rose, Shevlin, Winter, & O'Raw, 2010). However, this same legal framework exacerbated religious discrimination, it created a 'Catholics first' system (E. Daly, 2009) which has caused concern on the part of the UN Committee on the Elimination of Racial Discrimination in more than one occasion (Kilkelly, 2007).

Sometimes, EPA coverage concerns only some parts of the population. For instance, parental involvement in school matters in Ireland (6) has increased since the 1970s but it still at a lower level than desired (Eivers & Creaven, 2013; Mac Giolla Phádraig, 2010). The HSCL improved parental involvement in DEIS schools, but there is no information about non-DEIS schools (Inspectorate, 2011b; Weir et al., 2011).

There are limitations in the scope of this analysis. For instance, the list excludes overarching strategies and umbrella frameworks such as the National Anti-Poverty Strategy (Government of Ireland, 1997), the National Development Plans (Government of Ireland, 2000, 2007), the National Action Plan against racism (Department of Justice Equality and Law Reform, 2005) or the Strategy for Science, Technology and Innovation (Department of Enterprise Trade and Employment, 2006). Altogether, these strategies enabled policy development and created synergies between education policies and other reforms. This may have altered country alignment with the EPA in ways that were not fully captured by Table 5.2.

Nevertheless, some general comments on Irish policy convergence towards the OECD "model"—the inverted quotes are necessary, given the scarce coherence highlighted in Chapter 3—are possible. A country disregarding the EPA would have a predominantly white table, whereas a country strongly aligned with it, both in intentions and in practice, would have a green table. The table for Ireland has two features: it is mostly blue, suggesting good policy alignment with the EPA (though there is hardly any intervention specifically trying to decrease streaming); and it has become greener in recent years and especially since 2005. This was the effect of DEIS, the free pre-primary school year and Project Maths, which have convincingly managed to introduce some key OECD recommendations in the Irish system.

The impact of EPA-aligned policies on achievement and equity

Of the many policies theoretically in line with OECD recommendations, only four have been thoroughly and effectively implemented: centralised external exit examinations (CEEEs), DEIS, the free pre-primary school year and Project Maths. The latter two are too recent to be evaluated, but there are some indications that Project Maths is on the right track to produce systemic changes. CEEEs were already in place before PISA. DEIS, instead, was implemented in 2005 and there is evidence that it is having an impact on national assessments. One could tentatively claim that, in this circumstance, introducing a policy aligned with the EPA is having positive results on the system. Note that the importance of equity finds support in educational literature: perhaps, introducing a policy aligned with the EPA was beneficial because the EPA itself is aligned with research.

According to the OECD argument, PISA should be able to detect changes in the performance and equity levels of Ireland as a result of its policies. This argument can be put to the test using Willms' (2006) LSA-based policy evaluation framework. The framework states that the effects of 'universal' interventions should be captured by changes in mean country outcomes, 'SES-targeted' interventions by the Strength and Slope of the socio-

economic gradient, and that 'performance-targeted' interventions should lower the percentage of students scoring below proficiency levels 1 or 2 (Table 5.3).

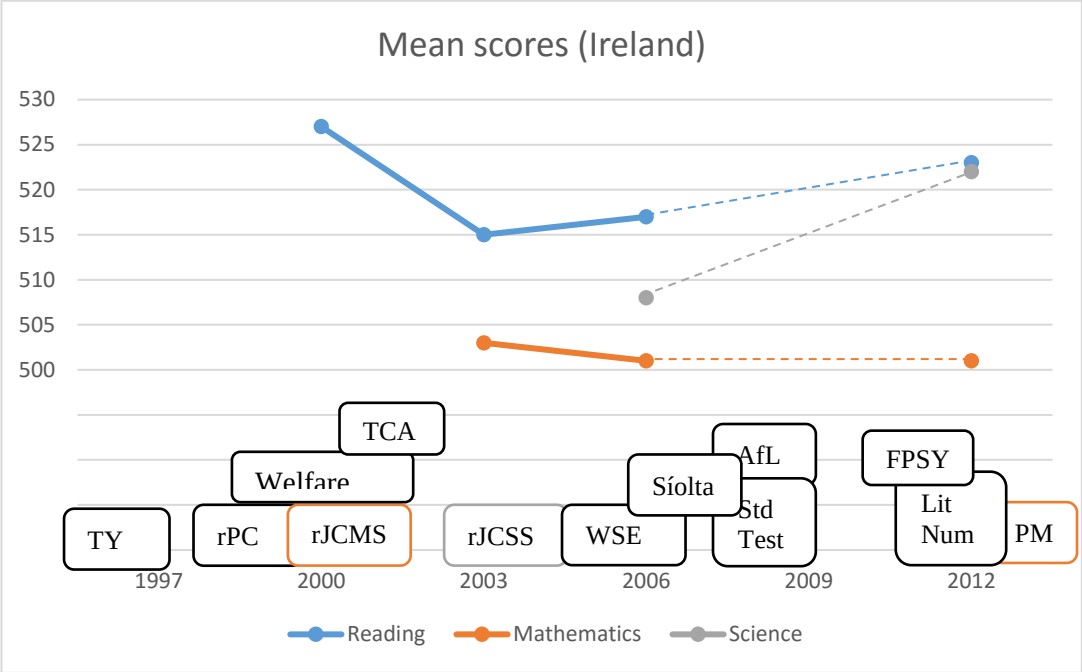
Table 5.3: a framework to detect the effect of Irish policies on PISA outcomes

Type and variable	Example policies in Ireland	Abbreviation
Universal interventions *strive to increase the educational performance of all children through reforms that are applied equally across the schooling system.* ***Variable:*** *mean country performance*	Transition Year revised primary curriculum Education (Welfare) Act No. 22 Teaching Council Act revised Junior Cert. Mathematics Syllabus revised Junior Cert. Science Syllabus Assessment for Learning Síolta Quality Framework Whole School Evaluation / self-evaluation Standardised testing Free pre-school year National Literacy and Numeracy Strategy Project Maths	TY rPC Welfare TCA rJCMS rJCSS AfL Síolta WSE Std Test FPSY Lit Num PM
SES-targeted interventions *aim to improve the educational performance of students with low socioeconomic status by providing a specialised curriculum or additional instructional resources.* ***Variable:*** *Strength and Slope of the socio-economic gradient*	Home-School-Community Liaison scheme Early start Breaking the Cycle National Anti-Poverty Strategy Giving Children an Even Break School Completion Programme Delivering Equality of opportunity In Schools	HSCL Early BtC NAPS GCEB SCP DEIS
Performance-targeted interventions *provide a specialised curriculum or additional instructional resources for particular students based on their levels of academic performance.* ***Variable:*** *Percentage of pupils below level 2*	Junior Certificate Support Programme Learning Support Reading Recovery Delivering Equality of opportunity In Schools National Behaviour Support Service	JCSP LSupp RR DEIS NBSS

Source: adapted from OECD (2013e, pp. 108–109); Willms (2006)

Figure 5.1 is a timeline of universal interventions and country mean scores. Policies focusing on raising performance in a specific domain were colour-coded. Scores from 2009 are missing following evidence that PISA results for that cycle were unreliable (Cosgrove & Cartwright, 2014). Therefore, the trendline connecting 2006 with 2012 results is dashed.

Figure 5.1: Universal interventions and performance in Ireland



It can be noticed from the graph that, despite the "universal" nature of the interventions, trends in the three PISA domains are slightly diverging, but it is not clear why. There is extensive evidence, which could not be covered in this chapter, that the fundamental aspects of teacher training, induction and CPD have not been touched since the 1970s (Conway, Murphy, Rath, & Hall, 2009; Coolahan, 2007). As a result, it is possible that only the most effective policies, like those on science education, managed to affect student scores. Yet their effectiveness was questioned earlier in this chapter.

Figure 5.2 and Figure 5.3 show the Slope and the Strength of the socio-economic gradient in Ireland. In this case, 2009 results were used to produce these graphs under the assumption that the additional error affected all students equally. Notice that, as was the case in Chapter 4, the charts for the Slope and the Strength follow different patterns even though they are supposed to reflect to sides of the same equity measure. This suggests once more that the actual error of PISA outcomes may be largely underestimated by the OECD.

Figure 5.2: SES-targeted interventions and the Slope of the socio-economic gradient in Ireland

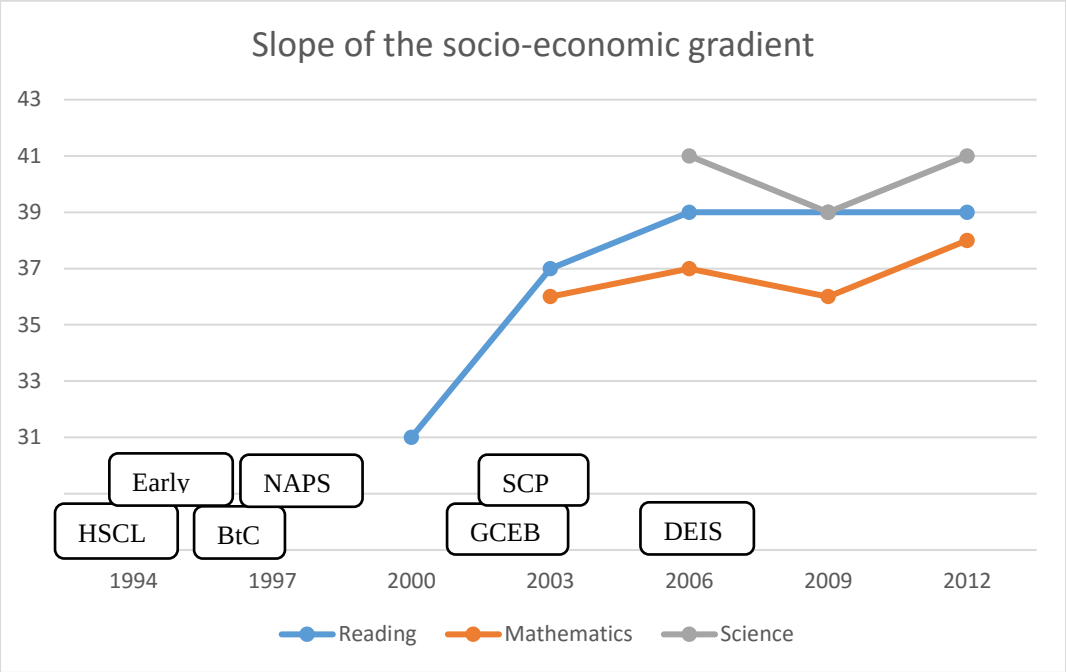
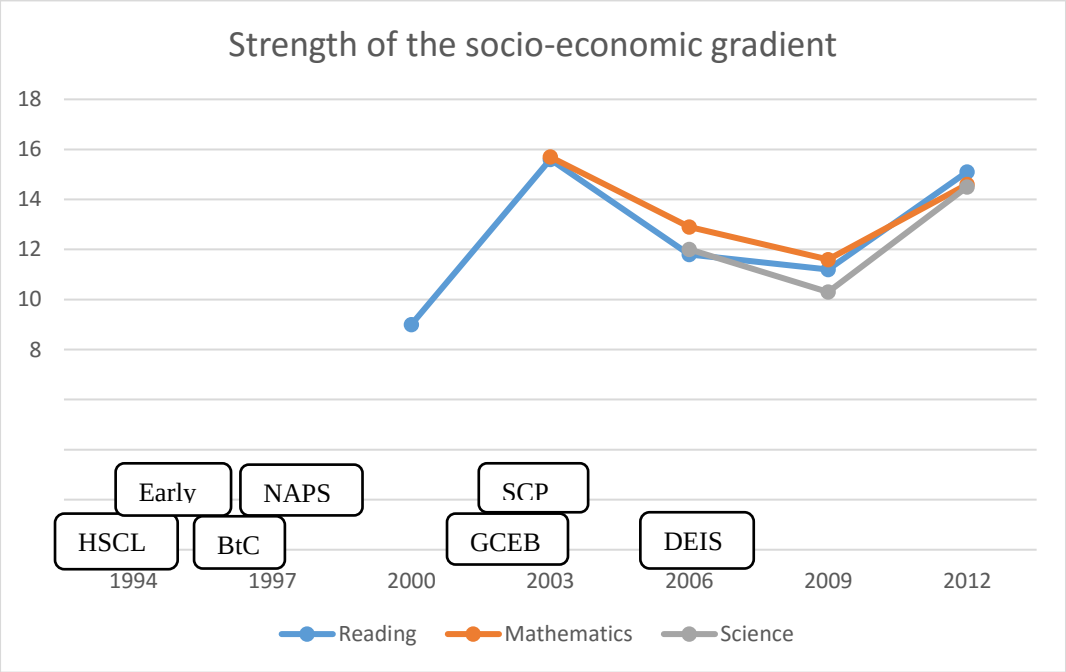


Figure 5.3: SES-targeted interventions and the Strength of the socio-economic gradient in Ireland



Inequality levels in Ireland are very close to the international average of about 40 PISA points for the Slope and 14% of the "explained" variance for the Strength. It is not clear why the socio-economic gradient for Reading was lower in 2000, but the trend has been rather stable since 2003, regardless of the various interventions and especially of DEIS.

Finally, Figure 5.4 and Table 5.4 attempt to capture the effects of performance-targeted interventions on the performance of the highest- and lowest-achieving students. Once again, values for 2009 could not be used.

Figure 5.4: Performance-targeted interventions and performance in Ireland

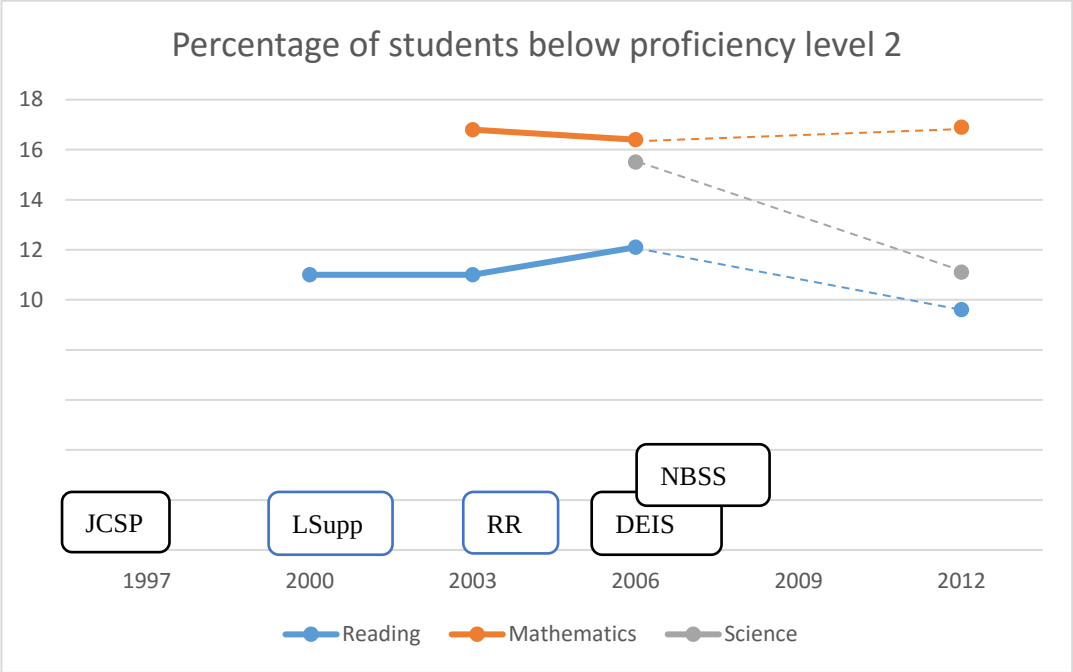


Table 5.4: Changes in the performance of the highest and lowest achievers in Ireland between the first comparable assessment and PISA 2012

	Scores	**% < lv2**	**% > lv 5**	**10th score**	**90th score**
R (2012-2000)	NS	NS	-2.8	NS	NS
M (2012-2003)	NS	NS	NS	NS	NS
S (2012-2006)	+14	−4.4	NS	+18.6	NS

Source: PISA database and national reports for Ireland. "Scores" = change in country mean scores; "% < lv2" and "% > lv5" = change in the percentage of students performing below the performance level 2 or above level 5; "10th score" and "90th score" = change in the scores of the 10th or the 90th student percentiles; "NS" = Non-Significant)

Table 5.4 shows that significant changes were limited almost exclusively to the Science domain and driven by the score increase of the low achievers. This result is puzzling: if it captured the effect of the science curriculum, it is unclear why only low achievers benefitted from it. Moreover, most policies targeting lower-performing students did not focus on Science, but on Reading and, to a much more limited extent, Mathematics. Nevertheless, the only significant change in either domain was a decrease in high achievers in Reading.

Overall, Willms' framework did not help to capture important relationships. This is because categorising policies by outcome does not take into consideration the fact that policies are qualitatively different. Interventions within the same group may touch deeper or more superficial components of the education system. Some affect 15-year-olds more directly than others. Structural policies such as delaying streaming by one year can have an immediate and strong impact on PISA scores (as was the case in Poland), whereas changing pedagogical methods is likely to have a slow and cumulative effect (as seems to have been the most common case according to evidence from Chapter 4).

Conclusions to Chapter 5

This chapter analysed policy alignment with OECD recommendations in Ireland and the extent to which EPA-aligned policies affected PISA outcomes using data collected through documentary analyses and expert interviews. The alignment of each policy was considered along the acknowledgment—enactment axis, whereas the concept of delta convergence was borrowed to interpret alignment over time. Policy implementation was accounted for, and the plausible impact of national reforms was explored using Willms' (2006) framework.

Evidence of country alignment with OECD recommendations

Hopkins, Pennock, & Ritzen (2008) reported that PISA 'achieved relatively low levels of impact on policy formation' in Ireland (p. 36). The expert responding to the e-mail questionnaire for this research, in contrast, stated that Irish reforms had been highly in line with the EPA. Both claims may have some truth: policy alignment of Irish policies with the EPA was already high before 2009; on the other hand, it was with the apparent drop of results in 2009 (after Hopkins et al.'s visit) that PISA became a stand-alone education topic among Irish policymakers.

Historically, Ireland has been very attentive to its international dimension, whether this meant following European or OECD practice, or participating to LSAs. Because of this, the OECD visit of 1991 had a spillover effect on the national legislative framework for education, the Education Act (1998). Changing existing practices took varying amounts of time.

In 1991, the OECD inspectors observed widespread frontal teaching, overreliance on textbooks and scarce attention to mathematics and science. Similar findings were reported in the IAEP2 and TIMSS 1995. However, what constituted a 'wake-up call' for science was not so for mathematics. Many science policies and initiatives emerged between 1999 and 2004, and there is some evidence of their adoption in the classroom. Instead, mathematics instruction seemed more deeply ingrained in the system; one has to wait until the 2012 pilots of Project Maths to see strong changes in national practice, and even then they were met with a certain resistance.

Similarly, reforms of the Inspectorate beginning in 2005 reflected a request for stronger data-based school accountability, but it is not clear whether the increased bureaucratic burden has effectively altered the relationship between the schools and the government. The teacher continuum (training, induction and CPD) also looks very similar to how it was in the 1970s, in spite of the 1991 OECD inspection and many years of debates and small amendments.

Policies like standardised testing in primary schools and the publication of inspection reports appear to be responding to international demands, but it is not evident that they are

influenced by the *educational* demands of the OECD—more by a national reinterpretation of new public management. Therefore, it is difficult to gauge the extent to which they can be said to be "aligned" with the EPA.

The 2011 National Literacy and Numeracy Strategy is the best example of policy enactment; it was developed in response of PISA outcomes, it draws from the PISA curriculum and literacies, and it will be evaluated by PISA. And yet, calls for a stronger focus on literacy and numeracy date back to the mid-1990s, so PISA may have just been the catalyst for policy demands that had been circulating in Ireland for some time.

Discourses on equity entered Irish policymaking in the 1980s and have become a cross-party staple of the political agenda. The conception of equity in Ireland has traditionally been highly aligned with the OECD's. Many policies addressing socio-economic disadvantage have emerged over the years and most of them flowed into DEIS, a cohesive framework covering much of the EPA. There have also been legislative developments to cater for SEN and immigrant students, but the system is still characterised by religious and gender segregation.

At the national level, PISA is administered by the ERC. There is generally good alignment between the views of the OECD and those of the ERC, but they not always have the same priorities. To improve Reading scores, the ERC recommended both policies targeting low-achievement and increased streaming. The centre was generally satisfied with national policies for science and attributed lower-than-expected outcomes to inadequate PISA coverage and lack of student motivation. On the other hand, it strongly argued for a change in the mathematics curriculum and used PISA results to support its case.

Overall, these findings confirm previous research on the differential ability of policy recommendations to penetrate a system depending on their timing and the extent to which they are shared by like-minded policy actors. They also highlight the role of national project managers as policy brokers and of research centres as intermediary agents which, depending on their unique combination of autonomy, personnel and leadership, can substantively impact on the national policy agenda. Instead, there is only limited support to the idea that Ireland was a passive recipient of globalising forces when it comes to accountability policies. Evaluation and assessment reforms were developed in-house and according to the "house rules", even though they may have been re-branded as cutting-edge international policies in some occasions.

Evidence of EPA-aligned policy effects

Very few policies in Ireland present three characteristics simultaneously: being aligned with the EPA; having been implemented effectively; and being sufficiently old that effects on student learning can be evaluated. DEIS is the best example, and evaluations suggest that it is

having an impact on student outcomes. Preliminary positive results can also be reported for Project Maths. If one takes a step back and categorises policy by supposed area of influence, however (as in the impact section above), no clear impact of national interventions on PISA outcomes can be detected. Given that this is seen even when there is independent evidence of policy effectiveness, this is more likely a shortcoming of Willms' (2006) evaluation framework or of the ability of PISA to capture small policy effects.

Chapter 6: The use and impact of OECD education policy advice in the French Community of Belgium

This chapter examines policy alignment with OECD recommendations and the impact of this in the French Community of Belgium. Belgium has been a federal state since 1970, when the 1831 Constitution was amended to divide the nation in three Communities: French (*Communauté française de Belgique*, CFB[22]), Flemish and German-speaking. Each has a parliament and a government and, since 1988, each is in charge of education within its area. A comparison between the education system in the CFB and in Ireland is provided in the Appendix. As in Chapter 5, the chapter is articulated in the following parts:

- A historical overview of the main educational issues in the country before its first PISA participation, including country relations with the OECD, participation in other LSAs and the key educational legislative framework.
- A summary of the targeted EPA received after each PISA cycle, drawing from the national version of the report.
- A detailed analysis of selected policies, covering the context surrounding the policy interventions, their degree of alignment with the EPA and, when possible, any evidence of implementation and impact.
- A summary of country trends in policy alignment with the EPA.
- An evaluation of the plausible impact of national reforms on PISA outcomes.

In the case of the CFB, it is argued that PISA entered the political discourse representing "international standards" to which Belgian policies aspired to conform. Because of the high level of socio-economic and educational segregation of the Community, however, attempts to standardise educational structures, processes, pedagogies and curricula have been met with scepticism and resistance. Performance in Reading seems to be on the rise, perhaps thanks to the increasing student familiarity with PISA and to support mechanisms that were put in place to help low performers. However, EPA-aligned policies addressing systemic inequalities have produced little improvement in this field thus far, whereas grade repetition appears to be worsening.

[22] In 2011, the French-speaking parliament (*Parlement de la Communauté française*) voted to refer to the CFB as *Fédération Wallonie-Bruxelles* in all communications and publications. The move was contested by the Flemish Community, which rejects the name on constitutional grounds. The use of *Fédération Wallonie-Bruxelles* is widespread in the French Community, but since it is still the object of controversy, this study adopts the original *Communauté française de Belgique* (CFB).

Historical overview

The evolution of education between identity and segregation

A key concept to understanding the Belgian education system is "freedom of education" as sanctioned in Art. 24 of the 1831/1994 Constitution. "Freedom of education" has two meanings: freedom *to organise* and freedom *to choose* education.

Provided they comply with some standard regulations, any physical or legal person can establish and manage a school, including its timetable, resources and staffing, curriculum and pedagogical approach. The authorities controlling a school are called *pouvoirs organisateurs* (PO, organising powers). Since the transfer of most education matters to the Communities in 1988, the POs are organised into *réseaux* (networks) along overlapping dimensions: one can speak of *réseau officiel* (public) and *réseau libre* (independent and private); or *réseau confessionnel* and *non-confessionnel* (religious and secular networks). Another key distinction can be made between schools that are completely managed and financed by the CFB and grant-aided schools (both public and private). Almost all schools below the tertiary level receive some sort of public funding (Eurydice, 2016a). Table 6.1 summarises the different networks that schools and POs can belong to.

Table 6.1: The different networks organising education in the CFB

Public		Private	
Fully financed	Grant-aided		
Secular			Religious

The freedom to choose education acknowledges the historical chasm between the religious-conservative half of society and its non-religious and progressive counterpart: 'Belgian society is organised around big divides—linguistic divides, but there is also one which is typical of 19th century and is the relationship between the state and religion. […] It was decided that different societies would coexist in Belgium, and so […] you have a Secular and a Christian society' (B1[23]). The *Pacte Scolaire* of 1959 (L. 29-05-1959; M.B. 19-06-1959[24]) formalised this divide by granting parents the power to choose whether to send their

[23] Interviews with Belgian experts are coded "B" (for "Belgium") plus a sequential number. In some occasions, it was felt that even this system may compromise the anonymity of the participant, in which case the number was replaced with a bar ("B-"). All interviews were done in French and translations were done by this author except where noted.

[24] Following the Belgian convention, legislation is referred to with two dates: when it was passed by the parliament and when it was published in the official gazette, the *Moniteur Belge*. All laws and decrees cited in this chapter can be accessed from the *Gallilex* website (http://www.gallilex.cfwb.be/fr/, accessed 16 April 2016).

children to a secular or a religious institution. Today, the choice is played more on the axis of academic reputation (Ryelandt, 2013) but is still reminiscent of its historical origins.

The resulting system is a "quasi-market", with privately managed schools competing for students and receiving public funding based on a very specific idea of state neutrality (defined in D. 31-03-1994; M.B. 18-06-1994): for instance, any 'social advantage' a state school provides to its population (such as food, clothes, trips, sport equipment…) must be granted to all private schools in the area (D. 07-06-2001; M.B. 26-06-2001).

While criticised because of its segregating effects, the quasi-market in the CFB was not the result of a mercantile view of education as was the case in Ireland, but rather of a compromise ('*un compromis à la Belge*', B1) to ensure that parents could choose between secular and religious institutions when neither philosophy 'managed to impose itself in the education sphere (or anywhere else)' (Vandenberghe, 2004, p. 242)[25].

The history of education in the CFB is intertwined with the history of the Community itself and of the Walloon region. The relationship between regions and cultural-linguistic communities was complex. Identity issues crossed financial ones. The 1970s and 1980s in particular were years of substantive deregulation and intense political action.

Social services experienced heavy cuts under the restructuring of the 1970s, and many hoped that the upcoming transfer of powers of educational matters from the central state to the CFB would also entail a transfer of capital. When competencies were handed over between 1988 and 1989 however, budget allocations were frozen. This led to waves of teacher strikes including violent protests like one on 27 September 1990 that left many injured (Destatte, 1995; Guidet & Bouillon, 1990).

One attempt to solve the crisis was the establishment in 1990 of the *Conseil de l'éducation et de la formation* (CEF), a council bringing together representatives from all stakeholders to advise the government on education matters (Ryelandt, 2010). The purpose was to provide a platform where teacher unions could express their discontent without bringing it to the streets. The protests subsided, and not for long, only between 1992 and 1994, when two agreements between the leading political parties and the constitutional revision unlocked some funding for education (Centre de Recherche et d'Information Socio-Politique, 2015; Destatte, 1995).

The search of the CFB for structural and financial emancipation from the central government accrued the autonomist pushes of the *réseau libre* and the segregation within the education system. This conflicted with ideas such as whole-system evaluation and accountability, which started to take hold in those years in Anglo-Saxon countries and progressively seeped into the policy arena of the CFB.

[25] Almost all documents used for this section were in French; the translations were done by this author.

International relations: the OECD

The oil crisis of the late 1970s affected both educational resources and discourses until the turn of the century. Austerity policies were carried out by left- and right-winged ministers and gave rise to teacher protests in 1982, 1990 and 1996 (Grootaers, 2005)—'the last strikes we had of that amplitude' (B5). The 'quest for efficiency' (Beckers, Florkin, & Leroy, 2008, p. 11) and accountability increased the significance of student assessment and systemic evaluation. Schools were criticised because they were (apparently) unable to provide young people with the skills employers sought, while international results from IEA studies in the early 1990s became the object of 'a special attention of authorities wanting to impose restrictions' (*ibid*.).

Austerity policies were 'somewhat legitimised by several audits commissioned by Belgian policymakers to national and international experts' (Grootaers, 2005, p. 32). These included a McKinsey report in 1986 and an OECD review in 1993 pointing to the higher cost per student in Belgium compared to other countries. It has been argued (Grootaers, 2005; also B2) that contacts with the OECD, including participation in the INES project, provided Belgian policymakers not only with the evidence, but also the right narrative to justify cuts in public spending.

Consistent with its role of knowledge mediator handing out both problems and their solutions (Chapter 1), the OECD (1993) pointed to the shortcomings of the system and explained that resources should be allocated on a need-basis rather than distributed to everyone equally. Note, however, that the report was requested by the three MoEs of the French, Flemish and German-speaking communities (Ryelandt, 2010, p. 34)—an example of voluntary submission to peer reviewing (Chapter 1) and a reminder that national policymakers are not simply passive recipients of globalising forces.

The background to the review and the national report (Ministère de l'Education, 1991; OECD, 1993) flagged up a lack of clear objectives and governance mechanisms in the education system (CEF, 1992b). The recently-established CEF was immediately receptive to the advice and published two policy documents mostly in line with the OECD recommendations (CEF, 1992a, 1992b). Since then, the CEF has often played the role of intermediary forum (Draelants, 2007)—a broker charged with translating international recommendations for a national audience (Cattonar, 2010).

On the government's side, proposals for the introduction of a steering committee to develop educational strategies began to surface in 1993 'inspired' by the OECD report (Ryelandt, 2010, p. 39). A public hearing on education (*Assises de l'enseignement*) was called and 'took place in Charleroi on 13—14 May 1995 to examine the general issue of education

and its financing' (Beckers & Voos, 2003, p. 4). The outcomes of the consultation informed the contents of the key educational framework for the CFB, the *Décret « Missions »* of 1997.

Key educational framework

Drawing from OECD-informed CEF advice (1992b, 1994b) and from the *Assises de l'enseignement* of 1995, the first proposal of the *Décret « Missions »* was submitted by the Green Party 'to fill the gap' represented by the lack of educational objectives and evaluative instruments in the CFB (Cheron, 1995, p. 2; also in Dupont, Neven, & Deghilaghe, 1997, p. 18). When the *Décret* was eventually passed (D. 24-7-1997, M.B. 23-9-1997), it provided the legal and conceptual framework for most education policies of the following 15 years, including the introduction of the Core and the Final Skills (*Socles de compétences* and *compétences terminales*) that students should acquire and which are common to all networks; or the establishment of a Steering Committee in 2002 for system monitoring and evaluation.

Table 6.2 summarises the alignment between the *Décret « Missions »* and the EPA.

Table 6.2: Contact points between the Décret « Missions » *and the OECD education policy advice*

Equality of opportunities to learn
▪ All actors should 'ensure equal access to all types of education for girls and boys' (Art. 10) ▪ One of the four objectives of education is to 'ensure to all pupils equal chances for social emancipation' (Art. 6) ▪ Each school is required to establish in its education plan how it will cater for special education needs (Art. 67)
Quality of provision
▪ Collaboration among schools is promoted as a means to ease transition between primary and secondary education (Art. 14) ▪ In each school, teachers, students, parents and other education professionals participate to the drafting of the school education plan (Art. 69) ▪ Schools are requested to produce a yearly 'activity report' (Arts. 72–73). In the first version of the decree, the report was sent to the Steering Committee and was highly confidential. Mentions of confidentiality were removed from subsequent amendments, and the report became one of the sources of data available to the Inspectorate. ▪ CPD providers receive sample Core Skills assessments for teacher training purposes (Art. 19), and information about the provision of CPD must be included in the school activity report (Art. 73)
Student performance
▪ Introduction of standardised educational outcomes. Note that the Core Skills 'give priority to reading for understanding, writing and oral communication, as well as the command of basic mathematical tools with a view to problem-solving' (Art. 16, § 3), but also to one modern language. ▪ A commission to develop and disseminate standardised tests of the Core Skills (which are then delivered to CPD providers) is also established, but these tests are sample materials and are not meant to be used in the classroom for assessment purposes (Art. 19)

Source: *Décret « Missions »*, author's translation.

There are some similarities between the Education Act (1998) in Ireland and the *Décret « Missions »*, like the idea of "gender equality" and an attention to special education needs (though SEN legislation in the CFB was not formalised until 2004). Irish policymakers

developed the concept of "educational disadvantage" in more detail, whereas Belgian policymakers were more attentive to stakeholders' involvement and cooperation—perhaps because of their multi-layered and multifaceted education system.

An interesting measure in the *Décret* is the supplementary year (*année complémentaire* or grade 1S). As advocated by the OECD (1993), schools should enable each pupil to progress 'at his or her own pace' through AfL and differentiation (*Décret « Missions »*, Art. 15). Therefore, a supplementary year 'adapted to his or her learning needs' was established for those pupils who were 'led' or 'brought' to complete ('*amené à parcourir*') primary education or the *premier degré* of lower secondary in one extra year.

In its intention, the supplementary year was seen as a fairer intervention than grade repetition to address underachievement—since grade repetition is often used simply as a disciplinary measure (Baye, Chenu, Crahay, Lafontaine, & Monseur, 2014). In practice, however, the *année complémentaire* was modelled on the existing system whereby schools enjoyed substantial freedom to decide who should or should not go through an extra year in education (hence the wording about pupils *being brought* to complete education). As a result, the supplementary year became just another form of grade retention, which is largely recognised to be a harmful and expensive policy (Education Endowment Foundation, 2016; Hattie, 2008).

Previous participation in international assessments

Just before PISA, the CFB participated in TIMSS 1995. The greatest issue was school participation rate, which fell below 85% because of teacher resistance (e.g., Beaton, Mullis, et al., 1996, p. 21). Positive comments were expressed with regard to teacher opportunities to meet and collaborate, whereas the issue of grade repetition was flagged up. TIMSS 1995 did not receive media attention, and this was the main difference between TIMSS and PISA: their findings were similar, but they reached the public only through PISA (B4).

Student results in TIMSS Mathematics were 'encouraging' because the CFB was 'slightly above the international average' (Monseur, 2000, pp. 16–17), but the performance advantage of the Flemish community on the CFB was 'more than a [school] year' (p. 17). A CEF analysis of the report suggested that gaps in teacher knowledge resulted in instruction that overemphasised procedural knowledge to the detriment of problem-solving (1998b). TIMSS 1995 was one of the many information sources that contributed to a wider debate about the purposes and instruction of mathematics in the CFB.

A second analysis (CEF, 1998a) was produced for Science. In this case, the CEF noted 'unsatisfactory results in a domain where skills are rapidly-evolving and an integral part of everyday life' (p. 4). The CEF recommended developing scientific literacy starting in primary

education by focusing on the relationship between science and society, problem-solving and experimentation. It also asked for a 'rigorous external evaluation [… to] measure in particular the effects of the actions taken, without waiting for the next IEA surveys' (p. 5). It is worth noting that, unlike Ireland, the IEA drew the attention towards teacher under-reliance (rather than over-reliance) on textbooks. This was deemed a problem by the CEF (1998a), and as a result, while other countries were trying to limit teacher use of textbooks because of concerns over their effectiveness, the CFB passed a decree (19-05-2006; M.B. 11-08-2006) to increase it among the scepticism (and criticism) of many experts (B7).

TIMSS participation served as the background for an interesting political exchange on 21st May 1997 between Laurette Onkelinx (Socialist Party, centre-left, MoE 1995–1999) and Pierre Hazette (Liberal Reformist Party, centre-right, MoE 1999–2004; see Parlement de la Communauté Française, 1997). The conversation aptly captures the complex relationship between "international" recommendations and national policies.

Hazette accused Onkelinx of presenting a draft of the *Décret « Missions »* that was too lenient towards students, because it removed barriers such as grade retention or exit examinations. Hazette thought that these measures were necessary 'to increase the pressure on students so that they could provide quality work' (p.17). He backed his argument with results from reading and mathematics assessments of 'the University of Liège' (*ibid*.)—in fact, the university-administered IEA RLS and TIMSS. Onkelinx defended her choice of having 'semi-external examinations' (*ibid*.) that would be internally developed by schools following governmental blueprints. She also added that low science results in TIMSS could be explained 'on the one hand, in part by the assessment methodology of the OECD [*sic*] for this subject and, on the other hand, by the perverted effects of [the] education system, particularly in science education' (p. 18).

Two features in this exchange are worth noting. Firstly, neither politician was completely in line with current EPA: Hazette supported both centralised examination and streaming policies, whereas Onkelinx was in favour of local autonomy but did not mention any accountability measure. Secondly, both seemed to be hardly aware of the data source. Hazette attributed TIMSS ownership to the University of Liège whereas Onkelinx to the OECD. Can one claim that *the IEA* was influencing national policymaking in this instance, when the product was being used but the owner was misstated?

Onkelinx's mistake may be evidence that the OECD was already being recognised as an LSA provider, but her response also shows that she did not simply accept everything the "OECD" said: the CFB did not satisfy the guidelines for sample participation rates in TIMSS 1995, but Onkelinx waved this off as a methodological issue with the assessment, moving the fault from the country to the international organisation.

The education policy advice for the French Community of Belgium

PISA is administered in the CFB by an independent research centre attached to the University of Liège, the *Service d'Analyse des Systèmes et des Pratiques de l'Enseignement* (Service for the Analysis of Systems and Practices in Education, ASPE, http://www.aspe.ulg.ac.be). The ASPE was founded in the 1960s by Gilbert de Landsheere—a key figure at IEA—under the name 'Laboratory for experimental pedagogy'. It has traditionally focused on issues of justice and fairness in education.

As mentioned in Chapter 1, it is hard to refer to the OECD as an *international* player when it comes to PISA in the CFB. ASPE researchers were involved in the INES programme, and they reportedly had a key role in promoting PISA to the Directorate for International Relations at the government of the CFB (Cattonar et al., 2009). Today, ASPE staff are not only in charge of the national implementation; they are represented in the PISA Governing Board, the PISA consortium, the Technical Advisory Group, the Reading and the Questionnaire Expert Groups. At the same time, the centre has been involved in developing, evaluating or advising the government on national education policies for many years, leading some detractors to question its actual independence from political sides (Cattonar et al., 2009)[26].

The ASPE appeared more aware of its role for policy than the ERC. The staff argued that there is an expectation that PISA data will be used by national project managers (NPMs) to produce policy-relevant reports. At a meeting in which the current research was presented, some claimed that PISA is not meant to generate any kind of general or international policy advice (i.e., the EPA) at all, but only to provide the data for national research centres to deliver tailored policy recommendations.

A few comments can be made on this point. Firstly, the pivotal role of the ASPE in brokering PISA knowledge to the CFB is undoubted but not representative of all countries. For instance, developing countries often have to rely solely on international reports as they do not have the capacity to re-analyse PISA data (Best et al., 2013; Gilmore, 2005).

Secondly, it was shown in Chapter 3 that international reports provide policy advice which might not be directly connected to PISA findings. The implication is that while NPMs may be a convenient means to tailor the EPA to a country, the EPA itself would still exist without national research centres or PISA. It could also be argued that while NPMs are supposed to work with PISA data, there is probably an expectation from the OECD—the Secretariat in particular—that national recommendations will not be too dissimilar from those in the

[26] This author was informed that the ASPE is in fierce competition with the academic group at the Université Catholique de Louvain that produced the cited report (Cattonar et al., 2009). The ASPE strongly contests its findings or academic value, suggesting that they were politically motivated to discredit the centre.

international report. The Secretariat is not simply a data provider, it is a policy advisor with a certain degree of autonomy and an agenda: if advice were systematically ignored or contested, its political credibility would be affected.

Finally, there is hardly any difference for practical purposes between claiming 'the EPA is supposed to influence national policies' and 'the EPA is supposed to influence national policies *through the NPMs*', provided changes in the EPA due to the NPMs' contribution are also analysed, as is the case in this research. Following the structure of the previous chapter, this analysis is provided in Appendix 3 (Tables 6.A–D) and summarised below.

PISA 2000

There is a remarkable alignment between the international and the national versions of the PISA 2000 report. The OECD and the ASPE identified similar problems, in particular the high inequality level of the education system. In line with their pre-PISA work, the ASPE tended to stress the negative effects of grade repetition.

The ASPE noted that issues with reading literacy were known and due to 'a tendency to consider learning to read as concerning almost exclusively the first two years of primary education' (Lafontaine, 2001a, p. 43). Citing post-PISA inspection reports, the ASPE argued that current reading instruction did little to improve the students' literacy and recommended supporting students with reading difficulties and improving teacher training and professional development in this area. The criticism on grade repetition and the reading curriculum was seconded by the CEF (2002), which also noted that the government had 'only partly responded' (p. 3) to recommendations that had been laid out ten years earlier (CEF, 1992a, 1992b).

In later reports, the ASPE critiqued the mainstream pedagogical approach in the CFB because it rewarded uncritical memorisation and meant that most students could only 'execute routine operations' in the PISA domains (Lafontaine & Blondin, 2004, p. 68). On equity, 'in spite of structural measures aimed at compensating social inequalities [e.g., decrees introducing Priority Education Zones and positive discriminations …] the diagnosis is without appeal' (Lafontaine & Blondin, 2004, p. 86). More was asked from the government to tackle grade retention practices: 'Even though there have been several conferences and debates around PISA […] we have not yet moved on […] from the debate stage to actions and reform proposals' (Lafontaine et al., 2003, p. 197).

The Flemish Community stood out for its better performance and its traditional teaching methods. Lest teacher-centred education could be taken as a viable policy option by the government, Lafontaine (2002) cautioned that the arrow of causality could go from higher performance to teaching style, rather than the opposite. In other words, it was argued, the

better performance of Flemish students could be explained by better socio-economic conditions, higher parental engagement and a more 'academic' demeanour of students—not by better instruction: 'if Flemish teachers are more demanding, it is because they can "get away with it", since their pupils perform better and are therefore more receptive with this kind of push…' (Lafontaine, 2002, p. 225).

It is unclear whether the impression that Flemish students were more academically-oriented may have been influenced by pre-existing regional bias. An interviewee said:

> 'What I am going to tell you is not at all scientific […] but… what you hear quite often… is that– pupils in the Flemish Community are more… *serious-minded* in school than French-speaking pupils, and that they are more disciplined […]. People [in Brussels] often say that there is an *enormous* difference in the atmosphere around French-speaking schools, where all students, they arrive late, they make noises, they are everywhere, they play… and: The Flemish schools where all students arrive like little soldiers, all responsible and disciplined.' (B4)

PISA 2003

The ASPE was generally satisfied with Mathematics performance and devoted its attention to inequality levels (Baye, Demonty, Fagnant, Matoul, & Monseur, 2004a, 2004b). The outcomes of grade repeaters in vocational tracks and in disadvantaged schools were 'seriously worrying', and equity was 'the challenge to rise to' (Baye et al., 2004b, p. 9). It was shown in Chapter 3 ("The consistency of the EPA"), that correlations between equity and performance can be read in different ways, including the possibility that inequality does not prevent achievement. For instance, the inequitable Flemish Community was also high-performing. The ASPE, however, adhered to the OECD interpretation: 'the examples of Australia, Canada [etc.] show that it is completely feasible to reconcile effectiveness with equity […]. It is one of the main teachings of PISA' (Baye et al., 2004b, p. 10).

The visions of the OECD and the ASPE were also in line with respect to the interpretation of the high correlation between domains. Where in the previous report the correlation of Mathematics and Science with Reading produced a cautionary note that PISA literacies might not completely capture actual mathematical and scientific competencies but rather the ability to read long texts (Lafontaine et al., 2003), the 2003 report justified the verbosity of PISA as reflecting 'real-life' applications (Baye et al., 2004b).

The fight against segregation and the narrative of possibility became part of a number of publications targeting education stakeholders (Baye, Demonty, Fagnant, Matoul, & Monseur, 2005; Demonty & Fagnant, 2006), including teachers. Unlike the teachers' guides to PISA in Ireland, the "Document for the attention of mathematics teachers of the first and second stages of secondary education" (Commission de Pilotage, 2004b) did not simply reproduce

information from the national report: it suggested changes in classroom practice based on detailed analyses of released PISA items.

The document attempted to link the PISA vision to the Core and the Final Skills frameworks. It conceded that PISA did not completely reflect the mathematics syllabi of the different *réseaux*, but argued that this was a strength rather than a limitation. The problem-solving approach of PISA Mathematics, the document explained, could easily be integrated within the curricula of general and vocational tracks. The released PISA items would serve as reference material. The high proportion of low achievers was 'not acceptable […, therefore] *all mathematics courses should make room for teaching these kind of problems*' (Commission de Pilotage, 2004b, p. 48, emphasis in the text).

PISA 2006

The full national report was published two years after the international version (Baye et al., 2009). Once again, there was substantial alignment between the OECD and the ASPE, both in their understanding of PISA and in the EPA.

The ASPE commented positively on the Science results as evidence that the 'basic' science curriculum (comprising three or four science periods per week) 'achieves its objectives' (p. 113). Engagement, interest and motivation were strong points as well as viable policy options, because they were variables 'that the education system should be able to affect without having to overhaul its structures' (p. 150). There was a good level of gender equality which contrasted with the wide gap between science graduates, where the percentages were 30 to 70 in favour of males. A follow-up report (Lafontaine & Goffin, 2009) showed that gender expectations, which were often reinforced in schools, were at the source of the gender gap. In line with the EPA, the authors recommended changes in terms of orientation and guidance.

The hierarchical structure of the education system was accused of reducing the '"reservoir" of pupils at ease with scientific skills' (Baye et al., 2009, p. 222), and academic selection was indicated as a source of between-school segregation: '"good schools" have many students in the first stage and less and less in the following years. On the contrary, other schools have a structure resembling an inverted pyramid, taking in masses of pupils that were rejected elsewhere' (p. 232). It is 'one thing […] to enrol in a top school, a whole other thing to manage to stay there' (Monseur & Lafontaine, 2009, p. 163).

In this regard, the ASPE praised recent interventions attempting to address inequalities such as changes in enrolment procedures, in the Inspectorate, as well as 'remediation in primary education, a cap on student transfers between schools […], a reorganisation of the *premier degré différencié* and an attempt to make vocational streams more valuable', which

were seen as a ‘first step’ to contrast academic selectivity (Baye et al., 2009, p. 232). Some of these policies are analysed later on in this chapter.

As in 2003, the teachers’ guide to PISA 2006 suggested using published items for training purposes: ‘PISA questions can support learning whether the objective is to form scientifically literate citizens or […] to approach scientific problems from a wider perspective’ (Quittre, Baye, Fagnant, & Hindryckx, 2008, p. 45). In general, the document reads more like an endorsement than a critique of PISA, and the premises upon which the assessment is built were hardly questioned: ‘The objective of PISA is to assess whether young people are prepared to enter adult life’ (p. 3); ‘PISA adopts a “citizenship” perspective to assessment’ (p. 4).

PISA 2009

Unfortunately, only the summary report could be retrieved to inform this section (Baye, Demonty, Lafontaine, Matoul, & Monseur, 2010). Where PISA 2009 was highly problematic for Ireland, the CFB celebrated a 14-point increase in Reading, especially because the improvement was greater among lower-performing students (those under proficiency level 2, boys and immigrant students) and was matched by higher indexes for engagement and motivation—ASPE’s preferred policy lever. This led the ASPE to claim that ‘the general mobilisation of field personnel—teachers in particular—and of the system to try and redress a situation that was considered to be alarming in 2000 is starting to bear fruit’ (p. 16).

What the ASPE did not acknowledge was that some of the positive changes were not statistically significant: the mean score increase, the decrease in the share of lower-performing students and in between-school variance were not significantly different in 2009 from their 2000 values; the reduction in gender gap was significant, but the 18-point increase in boys’ scores was not (OECD, 2010d, pp. 191–198). Lack of statistical significance does not imply that changes were not real, but it is unclear why the ASPE decided to overlook significance tests for some variables. In other cases, findings from the international report (an improvement in teacher-student relations despite a worsening of classroom climate) were not followed-up, as though the centre were more receptive to certain issues than others.

Even though the optimism about improvements in Reading was not completely warranted, this was the message that reached policymakers, together with reminders about the still high inequality levels. The resulting text resembled OECD discourses in which praise (used to show that improvement is within reach) is moderated by cautionary notes (used to keep future evidence relevant): ‘we will have to remain vigilant and continue in the direction of narrowing the gap, lest we wake up groggy when the next results come out… in 2013’ (Baye et al., 2010, p. 17).

Analysis of selected policies

In this section, selected national policies are analysed following a tripartite structure: context and influences, points of contact with the EPA or the PISA curriculum and, when available, evidence of implementation and impact. As mentioned in Chapter 5, the first step contributes to an understanding of how "international" discourses are translated on a national level. The second step helps to differentiate the level of goals and values (acknowledgment) from that of alignment. The third step seeks to assess the effectiveness of EPA-aligned policies.

Several policies were analysed in this fashion (see Table 6.4 at the end of the chapter), but only some are discussed in this section: as in the case of Ireland, they are curricular reforms, policies on national evaluation and assessment, or interventions seeking to improve the equity of the education system. Evidence of alignment with the EPA and effectiveness from each policy feeds into the broader end-of-chapter conclusions which, in turn, contribute along with the cross-country evidence from Chapter 4 to answering the second and third research questions.

Changing curricula

Core Skills, 2001

Context and influences

The Core Skills (*Socles de compétences*)[27] are defined in the *Décret « missions »* as a 'framework of reference presenting the basic skills that should be practised until the end of the first eight years of compulsory education, and those that should be mastered […] as they are considered to be necessary for social inclusion and later education' (Art. 5).

There are competing theories about the goals and values of the Core Skills. The idea of "skill/competency-based education" first appeared in 1993 as part of the reform of the first

[27] The term "*compétence*" is not a perfect translation of "skill" in the sense of the word intended in PISA—a definition is never provided in PISA, it is another multi-vocal symbol (Turner, 1975). The *Décret « missions »* formalised "*compétence*" as the 'ability [*aptitude*] to use an organised combination of knowledge, know-hows and attitudes allowing to carry out a certain number of tasks' (Art. 5). In francophone literature (e.g., Dolz & Ollagnier, 2002; Jonnaert, Masciotra, Barrette, Morel, & Mane, 2007) the notion of "competence" involves the ability to use one's knowledge, resources and potential to address a real-life situation in practice. The competency-based approach in education is framed within socio-constructivism and is more akin to problem-based, situated or minimally-guided learning than with "skills for life".

At the same time, the French version of the framework (OCDE, 1999) translated 'knowledge and skills' alternatively as '*connaissances et aptitudes'* or '*connaissances et compétences'*. Consistently with the OECD and the most recent Belgian translation of the *Socles de compétences* (Ministère de la Communauté française, 2014), this research translates "*compétence*" alternatively with "competence/y" or "skill", but the reader is advised that this is for simplicity.

stage of lower secondary education (*1*er *degré*). In line with the OECD report (1993), the reform organised grade progression, envisioned measures to support struggling students and interpreted education in terms of skills rather than contents. Competencies were 'structured behaviours to attain a goal in a given situation' (Odrovic, 2003, p. 3).

To some, competency-based education was introduced to address equity issues; it was a way to 'remedy [...] school dropout, to answer […] to the desire to teach always more [content], and to react to a certain student disengagement' towards prescribed knowledge (Rey, Caffieaux, Defrance, & Marcoux, 2005, p. 3). Others thought that competency-based education was a 'political response to demands formulated by international organisations and the business world' (B2). In 1994, the director of the ASPE was summoned by the MoE Di Rupo, 'and Di Rupo apparently told him—I'll quote verbatim […]—"I am just coming from a meeting with the OECD, and now what they… they told us is what matters nowadays is to work on skills"' (B2). From another angle, the European Round Table of Industrialists [ERT] wanted education reforms that could increase the production of human capital to meet the demands of a rapidly changing world (Hogenboom, 2003).

The ERT policy preferences were highly aligned with the OECD's. Key points were 'a palette of skills covering both numeracy, literacy, critical judgement and a knowledge of the basics of […STEM], humanities, economics and social sciences' (ERT, 1995, p. 7); pre-primary and adult education; ICT; decentralisation paired with accountability; lifelong learning; and 'the practice of *benchmarking'*, which 'should be introduced to identify weak and strong points of education systems and to identify […] best practices' (*ibid*., emphasis in the text).

Economic interference in educational matters has always been resisted by teachers and 'certain pedagogical movements of the progressive left' (B4), and at the time there were claims 'that education had sold out to enterprises' (Beckers, 2007, p. 5). Some authors, including Beckers (2007), disagreed with this view. For instance, Romainville (2008) acknowledged the influence of the ERT and the OECD reports, but stressed that the impact from within the CFB of teacher complaints about student knowledge and competences was not negligible for the paradigm shift:

> 'This thesis about a shady conspiracy of the "business world" does not stand [up] to scrutiny. Rather, the competency-based approach was introduced in schools under the pressure of a number of interconnected internal and external factors, and it would be over simplistic to reduce this complex interconnection to the pressure of the economic and social stakeholders alone.' (Romainville, 2008, p. 36)

Contact points with PISA

Irrespective of their intended goals, it is possible to evaluate the alignment of the Core Skills with the PISA literacies. Attempts to link the two were found in the teachers' guides

(Commission de Pilotage, 2004b; Quittre et al., 2008), and other similarities can be detected by comparing the OECD and CFB definitions of reading, mathematics and science (Appendix 3, Table 6.E). The strongest connection is in mathematics, where the idea that problem-solving skills can be used to address real-life problems and become active citizens underpins both definitions. "Reading" is defined more narrowly in the CFB, and a connection with PISA Reading as a skill for life can only be made by considering the purposes of the whole French syllabus.

One interviewee (B7) argued that these contact points exist because the OECD and the Core Skills were inspired by the same pedagogical theories, which were brought into PISA by the expert groups. This backs the argument made at the beginning of this document whereby the "international" and the "national" levels are confounded in PISA; yet the alignment between OECD guidelines and country-level interpretation is never perfect, and some idiosyncrasies can be spotted.

For instance, despite the influence of Belgian experts on PISA literacies, there is hardly any alignment in the science domain. The objectivity of science is not taken for granted in the CFB, where a religious and a secular society coexist and school neutrality towards worldviews is regulated by specific legislation. As a result, questioning the role of science in society and recognising that scientific values are not immutable are important features of the science curriculum (e.g., CEF, 1998a). This can be contrasted with the Irish curriculum for science which, despite the Catholic heritage of the country, embraces instead the OECD definition of "scientific literacy" and stresses the value of the scientific process for personal and societal wellbeing (NCCA, 2008).

Nevertheless, it is interesting to note that science in Belgium, which unlike French and mathematics was not a priority subject in the *Décret « missions »*, has moved up a few places in the list after PISA and is now presented right after the other two Core Skills in official documents (Ministère de la Communauté française, 2014).

Implementation

The legislative procedure leading to the enactment of the Core Skills provides a good example of the extent to which freedom of education is a sensitive topic in the CFB.

The Core Skills were meant to come into force in 1999 (D. 26-04-1999, M.B. 27-08-1999), but three not-for-profit associations and some private individuals appealed against the government's decision to the Constitutional Court, claiming that the Core Skills would undermine their constitutional freedom to follow the pedagogy of Rudolf Steiner. The court accepted the appeal because of the backwash effect that the norm, which was deemed 'too attentive to curricular details', could have on instruction (Cour d'Arbitrage, 2001, § A.8.3). The court also mentioned that the decree was based 'on the standardisation of achievement

levels advocated for Belgium in the OECD report of 13 November 1998' (ibid.), but the report itself could not be retrieved in this study.

The government successfully circumvented the appeal, without substantially addressing the issue, by granting schools limited exemption from following the Core Skills provided they could prove this 'did not compromise the coherence of the education system' (D. 19-07-2001, M.B. 23-08-2001, Art. 10). Since then, the existence of a backwash effect of the Core Skills on teaching practice has been a contentious point. Some authors claim the Core Skills produce an 'invisible pedagogy' (Bernstein, 1975; in E. Mangez, 2004, p. 68). Others note that they are hardly prescriptive and that curricula and assessments remain highly decentralised (Romainville, 2008, also B2, B5).

In fact, while the language of the new paradigm was adopted across all *réseaux* (E. Mangez, 2004), the curricula of some smaller associations (e.g., the *Fédération des Etablissements Libres Subventionnés Indépendants*) were adapted to account for the Core Skills as late as 2008.

No rigorous study of the effects of the Core Skills on curricula and classroom practice is available to date, but Rey et al. (2005) and Sylla & De Vos (2006) argue that—in spite of the constructivist slant of teacher training programmes and the focus on competences of the Core Skills—lack of time, resources, teacher confidence and professional development resulted in teachers reverting to more traditional and easier-to-implement instructional methods.

Beckers (2007) noted that the quality of the Core Skills varied by subject. In another paper, she cited Inspectorate reports mentioning 'difficulties' (Beckers, 2011, p. 12) in the implementation of the skill-based approach. This was an understatement. The first Inspectorate publication reported 'major structural difficulties in [the] education system' (Service général de l'Inspection, 2009, p. 25), including weak internal coherence of the Core Skills and a lack of consistency in learning experiences even within the same school. Curricula and assessments were poorly aligned with the Core Skills.

In secondary education, two thirds of lessons were found to be at odds with skill acquisition, and while there was evidence of attempts at a different pedagogy, 'their practice was […] clumsy and intermittent' (Service général de l'Inspection, 2009, p. 12). Instruments devised to foster teacher cooperation and to help struggling students were seen as administrative burdens. One year later, the Inspectorate underlined again the inadequacy of the Core Skills as a basis for teaching and assessment: 'From one school to the next, from one classroom to the next, pupils do not learn the same thing and are not assessed in light of the same criteria, and this happens even in […] essential domains' (Service général de l'Inspection, 2010, p. 42).

Impact

After PISA 2009, Baye et al. (2010) argued that the reforms of the early 2000s, including the Core Skill regulations, had started to bear fruit in terms of student achievement. If this were the case, the evidence above suggests that Core Skills may have played an enabling role but not have had a direct effect. The framework 'has the great merit of existing' (Beckers, 2011, p. 11), because it helped to regulate the pedagogical freedom enjoyed by Belgian teachers. It was useful to shake up the system (Chenu, Crahay, & Lafontaine, 2014) but in everyday practice it suffered from clear shortcomings that other systemic interventions, such as the School Contract of 2005 (below), tried to address (Beckers, 2011).

School Contract, 2005

The *Contrat stratégique pour l'éducation* (Strategic Contract for Education; Gouvernement de la Communauté française, 2004, 2005b)—then renamed *Contrat pour l'école* (School Contract; Gouvernement de la Communauté française, 2005a)—was an attempt by the government and the MoE Arena to set a long-term strategy for education by involving all stakeholders in the deliberation process. While it was not strictly a curricular reform, it is analysed in this section because of its strong links with the Core Skills and its goal of raising all pupils' achievement.

Context and influences

Two factors provided the impetus for the School Contract. On the one hand, teacher consultations in 2003–2004 highlighted discontent among professionals about the reforms of the mid-late 1990s: teachers argued that they had been left out from the deliberation process, that new frameworks lacked clarity, and that the scarcity of resources and opportunities for professional development translated into difficulties in implementing the Core Skills in the classroom (Gouvernement de la Communauté française, 2004, 2005a, 2005b). On the other hand, PISA 2000 and 2003 had drawn attention to the mediocre performance of the CFB on an international stage and to its high inequality levels (*ibid*.).

PISA results showed that too many students did not achieve satisfactory performance levels in domains considered the bases for 'active citizenship'—which were therefore in need of greater attention 'without sacrificing other areas' (Gouvernement de la Communauté française, 2005a, p. 5; note the difference with the narrative for the National Literacy and Numeracy Strategy in Ireland, whereby other areas "had to be sacrificed"). Social and academic segregation, including grade repetition, were identified as the greatest culprits behind low achievement.

The School Contract was an attempt to raise both equity and performance simultaneously by working on two fronts. On the one hand, it provided a coherent conceptual and financial

framework for policy instruments that already existed (such as *Décret « missions »*, Core Skills, Steering Committee, legislation to contrast segregation) but were somewhat disconnected (de Commer, 2005). On the other hand, it introduced a new model in education governance where all stakeholders would be involved in the consultations.

The debates, compromises and negotiations lasted for months, and the resulting policy document provided a vision for education that was certainly all-encompassing, but not always coherent and in which many voices seem to overlap.

Education is '*the* priority in a democracy' (Gouvernement de la Communauté française, 2005b, p. 2, emphasis in the text), and education for citizenship fosters 'democratic values and intercultural dialogue', which are 'the fundaments of social cohesion' (2004, p. 1). Education enables individual and collective emancipation, and this in turn leads to (sustainable) economic growth and 'the well-being of future generations (2004, 2005b, p.1). 'Teaching evolves in a global context', years of austerity 'imposed' to the CFB severely affected society and the education system, and although school 'cannot and should not be accountable for all problems of society', measures could be taken on the 'quality, effectiveness and equity' of the system (2004, p. 2). Education is like a rising tide lifting all boats: 'for this, school should guarantee to everyone equal access to knowledge and know-how', which will enable young people to think for themselves and develop their personality (2005a, 1).

Contact points with the EPA

The School Contract had ten priorities and about 54 action points, and at least 37 of them (69%) had clear links with the EPA. They included: increasing the teaching time dedicated to French and mathematical literacy; developing supporting mechanisms for struggling students and reducing segregation; introducing standardised testing at different levels of education; de-gendering career paths and raising awareness among teachers about gender equity; developing a coherent policy for resource allocation and removing teacher allocation restrictions; making vocational education more appealing to all students; improving teacher training and professional development with a view to the quality, equity and effectiveness of the instruction; strengthening school leadership; improving teacher cooperation and parental involvement; strengthening the Inspectorate and the Steering Committee.

Implementation

A week after the publication of the School Contract, de Commer (2005) argued that it could succeed only if it managed to fit within other strategies of the federal government, which were outside the control of the CFB, and if teachers could be engaged. Work to address some of the priorities began immediately. The Steering Committee issued suggestions to improve teacher training (Commission de Pilotage, 2006), while the Universities of Mons and Liège collaborated on a project to tackle inequality within the 'ghetto-schools', so called because of

their high segregation levels (Demeuse et al., 2007; Friant, Derobertmasure, & Demeuse, 2008).

The Contract was supposed to undergo biannual evaluations, but 'nowhere' was there an interest in complying with these deadlines and the monitoring aspect was put aside (B4). Then, four years after the Contract, the government gladly reported that over 91% of the action points had been completed or were underway (Gouvernement de la Communauté française, 2009). The report mentioned unprecedented financing, great support for the reforms and concrete results but its findings were disputed by the opposition (CEF, 2015a).

In 2014, the new government commissioned a new evaluation from the CEF. Policy-wise, it was shown that a great number of decrees, frameworks, pedagogical interventions and instruments, research studies and CEF publications had seen the light of day between 2005 and 2014, though they were not matched by as many evaluations (CEF, 2015b). Some areas had been targeted by policy more than others; priorities such as raising attainment levels, improving links between employers and students or increasing systemic evaluation and monitoring had been pursued with greater determination than, for example, streamlining teacher qualifications. The CEF also noted the uneven distribution of policy enactment. Most interventions took place between 2006 and 2009, many changed substantively over the years, some appeared late and some (teacher training and CPD) were never attempted (CEF, 2015a, also B5).

Overall, the CEF argued that the School Contract managed to redesign the governance and administration of the education system: learning objectives were standardised and instruments for assessment, evaluation and monitoring were reinforced; the management of the educational structures was decentralised but schools were held more accountable; at the same time, more resources like teaching materials and equipment were unlocked. To summarise, where the OECD report of 1991 found an education system affected by severe lack of funding and with no central structures for steering and evaluation, the School Contract created one that had all these features (CEF, 2015a). Some interventions, however, 'collided more than others against the existing context' (p. 92). This was the case for new mechanisms of enrolment in lower secondary education and for changes in teacher training and interventions to support low achievers.

Impact

The School Contract had six objectives set for 2013: increase the attainment of all students; increase the performance of all students; reduce grade repetition; reduce school segregation and between-school variance; place vocational and academic tracks on an equal standing; reduce academic selection and streaming. The success of the second objective was to be

measured by an increase in PISA mean performance and a decrease in the percentage of low achievers.

Most criteria were not met: there was some improvement with regard to basic skill acquisition—particularly in relation with school completion—but it was confined to certain geographical areas and was still considered unsatisfactory; instead, the levels of grade repetition, school segregation and overall inequality had remained stable (CEF, 2015a). With respect specifically to PISA outcomes, the objective of reaching the OECD average was achieved only in Reading: mean Mathematics scores were already at OECD levels, and Science scores remained below the average. Similarly, a reduction in the proportion of very low achievers (below PISA level 1) was observed only in Reading, where the percentage decreased from 28.2 in 2000 to 19.2 in 2012 (*ibid.*). There were no differences in the size of the gap between the highest and the lowest performers (an equity measure, though not an explicit objective of the Contract), or between tracks.

These results led the CEF to argue that 'a system holding its actors accountable and regulating by performance evaluation is not sufficient to change what happens in the classroom' (CEF, 2015a, p. 92).

Changing evaluation and assessment

Assessment in primary education

Low-stakes external assessments in primary and secondary education were put in place in 1994 to monitor the education system as advocated by the OECD in 1991. Since they were positively received by teachers, the Inspectorate tried to exploit them to standardise teacher practice regarding the Core Skills, but the attempt was hampered by insufficient legal framing and institutional support (Lafontaine, 2001b). As in Chapter 5, however, this section will focus primarily on standardised testing of pupil progression, which should bear higher stakes for teachers and pupils.

Context and influences

A non-compulsory examination leading to a certificate of primary school completion had existed since the 1920s and another was introduced by the Catholic *réseau* in the 1960s. Participation varied by school network and area (C. Mangez, Maroy, Cattonar, Delvaux, & Mangez, 2009), and since exams were independently developed by school inspectors, the state network alone could count 10–15 different tests every year (B4). In principle, a unified assessment would have been a 'logical extension' of the *Décret « missions »* of 1997, the first policy document trying to turn a 'juxtaposition of schools' into a cohesive education system

(B3). But the *Décret* did not alter the fragmented state of educational certifications, because 'since there were no common programmes, there could be no common examinations' (B3).

The systemic inequality reported by PISA 2000, however, resonated with the MoE Nolet, who saw in a centralised examination leading to a common *Certificat d'Études de Base* (CEB, basic learning certificate) an instrument to ensure higher 'equity and, therefore, effectiveness' by addressing the segregation caused by streams and *réseaux* (Lafontaine, 2007, p. 6). Common standards (the Core Skills) were introduced in 2001, and in 2003 the Minister presented to the attention of the Steering Committee a 'request for suggestions regarding an increased regulation of first-level education in a quasi-market context' (de Commer, 2003; Lafontaine, 2007). The Steering Committee welcomed the reform provided that 'failing the test would not automatically entail failure to be awarded the certificate' (Commission de Pilotage, 2004a, p. 2).

The resulting decree (02-06-2006, M.B. 23-08-2006) strengthened the evaluative capabilities of the CFB in two ways: it made low-stakes external assessments compulsory and it introduced an examination for awarding the CEB, which became compulsory in 2009 (in 2011, an additional but non-compulsory examination was introduced at the end of the 1st stage of lower secondary). The low-stakes tests are triannual and at each cycle they focus on French, mathematics or one of the subjects listed in the *Décret « missions »*, while dedicating a 'particular attention' to technology and science (Art. 8 § 1). Similarly, the CEB is based on the Core Skills and has focused in recent years on French, mathematics and science or history/geography (Ministère de la Communauté française, 2015). French and mathematics were selected first because it was easier to agree on test contents and objectives (B6).

Contact points with the EPA

Both before and after the CEB examination became compulsory in 2009, the stakes have been high. Failure to attain the certificate means being unable to progress in secondary education at the same pace as other children. While the rest of the cohort enters the *1ère commune* (1C), a pupil without CEB is directed into the *1ère différenciée* (1D). At the end of 1D, he or she sits the examination again. If the CEB is not attained, the pupil continues in 2D; even if the CEB is attained, the pupil will not join his or her original cohort, because enrolment in the new academic year will be in 1C or in 1S (the supplementary year introduced by the *Décret « missions »*), not 2C (Ministère de la Communauté française, 2009b). In other words, this is a streaming policy that also entails automatic grade retention. In 2014, 11% of pupils did not attain the CEB at the end of primary education—a figure seven percentage points higher than 2013 and in line with 2009 results (CEF, 2015a).

The OECD endorses Wößmann's (2005) argument that centralised external exit examinations (CEEEs) can be used to standardise educational outcomes by holding schools

accountable, for instance by posting their students' results publicly. This may be true (see Chapter 3 for a critique), but the CEB is only superficially similar to the CEEEs described by Wößmann and the OECD.

The examination leading to the CEB is indeed centralised but it is placed at the exit of primary—not compulsory—education. The new system is considered fairer by some because it sets a bar at a common height (B7), but the test itself is not standardised and it is affected by reliability and administration issues compromising its fairness (Blondin & Giot, 2011, p. 77; also B4). In spite of the worries of school management regarding CEB-related accountability (Blondin & Giot, 2011), the decree excludes uses of data that may lead to "league tables" and the consequences for schools have thus far been very limited (see next section).

The new regulations were devised to 'reinforce equity in education by regulating qualifications by organising a centralised test' (Commission de Pilotage, 2004a, p. 1), and some argue that they did, combined with other decrees affecting the organisation of the first stage of lower secondary education (Lafontaine, 2014). As an individual policy, however, the CEB regulations did not meet OECD requirements of standardisation and accountability, even though they may have acknowledged the OECD goals and values in their intents. This is a reminder that the usefulness of a policy and the value of OECD recommendations (the focus of this thesis) are two separate aspects.

The Steering Committee and the Inspectorate

It has been argued that the establishment of the Steering Committee and the reform of the Inspectorate, along with other policies for systemic evaluation such as the CEB, signal a paradigm shift in the governance of education in the CFB from a model based on complete school autonomy to one seeking to centralise and standardise some of its components (C. Mangez et al., 2009).

The idea of systemic monitoring for steering purposes was already present in the CFB in the late 1980s, but it was the OECD report of 1993 that popularised the concept (C. Mangez et al., 2009). A 'light steering structure' was created in 1993. Its work introduced new ideas in evaluation such as external assessments (section above), the publication of the first education statistics at the Community level and suggestions about a possible reform of the Inspectorate. The body acquired more formal status in 1995, but the next few years represented something of a hiatus: The *Décret « missions »* of 1997 recognised the value of evaluation and monitoring but the steering modes it envisioned (non-compulsory tests and inter-network collaborations) were not in line with the view of the group. Subsequently, PISA 2000 results showing high inequality levels in the system 'accelerated the adoption of a new decree' that

strengthened and enlarged the existing steering group (Ministère de la Communauté française, 2008, p. 80).

The decree of 27th March 2002 (M.B. 17-05-2002) established the Steering Committee in its current form. The Committee was opened to representatives from parent associations and teacher unions and was charged with key responsibilities in system monitoring and evaluation, but also in strategy development, curricular standardisation and teacher training. It was a key player in the implementation of the School Contract and was involved with the ETNIC (a public agency for data and information management also created in 2002) in producing statistics on the education system. It receives data from external assessments, from CEB examinations and, since 2007, from the reformed Inspectorate.

In a draft version of the decree it had been suggested that the Steering Committee should develop PISA-based national indicators of equity, effectiveness and socio-economic conditions of the school population. The purpose was to 'assure the scientific nature of indicators and their effective coherence with international standards' in line with the national development plan (Charlier et al., 2002, p. 2), but the recommendation was not integrated into the enacted decree.

The reform of the Inspectorate (D. 08-03-2007, M.B. 05-06-2007 erratum 13-12-2007), on the other hand, changed three aspects in the existing evaluation system: it separated the evaluative and advisory functions within educational services and assigned inspectors to the former (as happened in Ireland with the White Paper of 1995); it relieved the Inspectorate from administrative burdens; and it extended its mandate to all school networks (C. Mangez et al., 2009). This was 'an enormous change for school inspections', as the focus shifted from individual teachers to schools (B5). The reformed Inspectorate claims to have been very active, producing around 3000 school reports per year, informing national debates, providing data to the Steering Committee and helping each school to improve (Service général de l'Inspection, 2012).

It is unclear, however, whether the new regime produced actual change in school practice. At least until 2012, there was no standardised measure or reference framework to evaluate teaching and learning, so the Inspectorate had to develop its own criteria and instruments (Service général de l'Inspection, 2012). It was found that the objectives, timings, modes and outcomes of inspections needed improving (*ibid.*). The legitimacy of inspectors was also called into question: the independent network felt bereft of the power to evaluate its own schools, and claimed that government inspections could not capture the quality of independent schools adequately (B1).

The government answered the criticism in the national development plan 2014–2019, where it set out its plans to 'stabilise' and transform the Inspectorate into a 'professional body

entrusted with the auditing of the quality of education in all schools as it happens already in other education systems' (Gouvernement de la Communauté française, 2014, p. 24). The independent network is currently more open to school inspections as long as it can retain autonomy in pedagogical matters (B5).

In terms of sanctions, they 'are light, very light on a collective level, and individual implications may be even less significant. The inspector produces a report which is transmitted to the authorities with requests for improvement […] and then either the teacher complies or there are other requests but no sanctions' (B1). As a result, 'not much' (repeated three times) happens after school inspections (B1).

Changing equity

Systemic inequality is, together with low performance, the most frequently-debated topic in the CFB, particularly after the OECD report of 1993 and 'the electroshock' of PISA 2000 (MoE Simonet, 2010, quoted in Grootaers, 2014, p. 5). It would be unfeasible to evaluate all interventions attempting to correct these trends; therefore, this section focuses on one specific policy regulating enrolment in lower secondary education as an exemplary case of how EPA alignment was reached but highly contested.

Enrolment regulations in lower secondary education, 2007–2010

Context and influences

The significance of new enrolment policies in the *1er degré* (the first stage of lower secondary education, see Appendix 2), which were introduced between 2007 and 2010, can only be understood within the wider context of a long-term fight of part of the society against socio-economic segregation and academic exclusion.

To contrast socio-economic segregation, Priority Education Zones (*Zones d'éducation prioritaire*, ZEPs) were introduced in 1989. Loosely based on the French model, the ZEPs were also similar to the Disadvantage Areas Scheme in Ireland (Chapter 5) because they unlocked additional resources for schools within specific areas (Friant, Demeuse, Nicaise, & Aubert-lotarski, 2008). The ZEPs were criticised because rich schools in poor areas benefitted by the targeting system for years, so the policy was eventually replaced in 1998 by a decree on positive discrimination.

Discourses on positive discrimination had appeared a few years earlier, when the CEF argued that 'the 1960s and 70s policy of equality of opportunities did not produce equality of outcomes. Rather, it benefitted those who were already advantaged' (CEF, 1994a, p. 3). Positive discrimination for resource allocation was seen as a way 'to give more to those who have less' (*ibid.*). A first hesitant attempt to translate the principle into practice was made by

the ‘schooling for success’ decree (14-03-1995; M.B. 17-08-1995). In 1996, the MoE Onkelinx called for ‘an ambitious policy on positive discrimination’ (quoted in CEF, 1998c, p. 5), which the *Décret « missions »* did not deliver. Thus, the CEF elaborated a policy proposal based, among other sources, on an OECD report on the French experiences in these matters (CEF, 1998c). The proposal was rapidly converted into a decree, which ‘aimed to ensure equal chances of social emancipation to all students, namely by implementing positive discrimination’ (D. 30-06-1998, M.B. 22-08-1998).

A socio-economic index was developed by a research team (Demeuse, 2000) and associated with each student; schools were ranked according to their average index and extra funding was unlocked above a certain threshold (Grootaers, 2005). Over the years, the decree was amended many times with changes predominantly focusing on modes of financing and calculation of the index: the most significant was the decree of 30th April 2009 (M.B. 09-07-2009), or *encadrement différencié* (differentiated management), which introduced progressive funding allocations to replace the binary system operating since 1998. There are, however, no clear evaluations of the outcomes of these policies (Friant, Demeuse, et al., 2008; Thilmany, 2011).

With respect to academic exclusion, in contrast, the issue had always been how to limit vertical and horizontal streaming (grade retention and tracking) while retaining the existing structure involving tracks and *réseaux*. The ‘schooling for success’ decree of 1995 attempted to reduce grade retention by introducing pedagogical cycles with automatic promotion within cycles. The *Décret « missions »* positioned the *1er degré* at the end of a pedagogical continuum terminating with the acquisition of the Core Skills. Since then, the *1er degré* was no longer ‘an open door’ to secondary education but ‘an airlock’ that could be exited only by those who managed to achieve the Core Skills (Grootaers, 2014, p. 10). This had repercussions on its relevance for policymaking and parental choices.

At the policy level, there was a need to account for pupil achievement in decisions concerning their progression. This translated into structural adjustments, including the introduction of the *année différenciée* mentioned earlier in this chapter, as well as measures such as individualised learning plans for struggling students (see D. 30-06-2006, M.B. 31-08-2006; D. 07-12-2007, M.B. 26-02-2008; or Ministère de la Communauté française, 2009a). From the parents’ perspective, the initial choice of a secondary school became critical to ensure that their children got through ‘the airlock’ without delays.

As can be inferred from this brief historical overview, the battle against socio-economic segregation and academic exclusion introduced new forms of risk for some stakeholders: advantaged families did not want to send their children to designated disadvantaged schools, neither did they want for them to end up in a differentiated stream. It is in this context that the

government tried to regulate school enrolment to ensure better chances to disadvantaged pupils. The outcomes are discussed below.

Contact points with the EPA and implementation issues

Enrolment in the *1er degré* had always been unregulated in the CFB: parents could choose any institute at any time (even years before the child qualified for entry), and schools had the power to accept or decline registrations with very few constraints (Ryelandt, 2013). A doubly-segregating mechanism was at work. On the one hand, those who were able to move within the system with the greatest ease could plan in advance the best learning path for their children and close doors to other children. On the other hand, since admissions were interview-based, personal relationships between schools and families were an asset: many schools regulated access by discouraging unwanted families to proceed with the enrolment (Draelants, 2014).

In 1997 and 2001, the *Décret « missions »* and inter-network financial agreements (*accords 'de la Saint-Polycarpe'*) attempted to streamline enrolment procedures by penalising social discrimination and obliging schools to accept all applications provided parents subscribed to the school ethos and pedagogy (Delvaux & Maroy, 2009; Ryelandt, 2013). This, however, did not solve issues such as early enrolment or the self-selection of disadvantaged families into disadvantaged schools (Ryelandt, 2013; B5; for a more general overview on self-segregation, see Hoelscher, Hayward, Ertl, & Dunbar- Goddet, 2008). A clearer policy direction was set by the PISA-motivated and EPA-aligned School Contract of 2005 (see above). One of the priorities of the MoE Arena was to 'say no to ghetto-schools' by targeting enrolment, student transfers and disciplinary exclusions (Delvaux & Maroy, 2009).

Enrolment and student transfer were addressed in the *Décret « inscriptions »* (D. 08-03-2007, M.B. 03-07-2007). The decree introduced two simple but ground-shaking procedures: it introduced a common yearly window for enrolment in the *1er degré* and prohibited student transfers for the first two years of secondary education. In other words, it precluded early access to prestigious schools by applying a 'first come, first served' policy and, in line with the EPA, it made it harder for schools to 'suggest' to struggling students that they should continue elsewhere (Ryelandt, 2013, p. 24). The political recoil that ensued cost the post to the MoE Arena. Teachers, parents, anti-heterogeneity ideologists, unionists and new lobbyists, journalists and politicians came from both sides of the political spectrum to accuse the Minister of attacking constitutional freedoms or merely to critique some side-effects of the reform, such as the few but highly-mediatised queues of people standing for days in front of the "best" schools to gain a place (Ryelandt, 2013). In early 2008, the MoE was 'promoted' to the federal government (Delvaux & Maroy, 2009, p. 8).

Her successor, MoE Dupont, tried to adjust the aim with a second decree, known as the *Décret « mixité sociale »* (D. 18-07-2008, M.B. 26-08-2008). The decree changed the

admission criterion from chronological to stratified-random: a percentage of places would be automatically offered to pupils coming from disadvantaged primary schools, whereas the remaining would be randomly allocated. Theoretically, this decree was even more aligned with the EPA than the former, as it reduced segregation by forcing each school to take in a certain number of disadvantaged students, rather than by "hoping" that poorer parents would be at the front of the queue (besides, it had emerged that most people queuing were socio-economically advantaged parents that had lost their early-reserved place).

Two main issues affected this second decree. Firstly, since the government could not agree on a unique randomised allocation procedure, schools were asked to devise their own method within certain constraints, causing logistic and administrative repercussions that lasted for months. Secondly, the decree suffered from what was renamed 'the enrolment bubble': strategically-minded parents sent applications to many schools to maximise their chances of being assigned a place, with the result that twice as many schools compared to the previous academic year figured as fully-subscribed (Delvaux & Maroy, 2009; Ryelandt, 2013). The decree was amended three times between 2009 and 2010 to no avail.

In 2010, a third major attempt to regulate enrolment was made with the decree 18-03-2010 (M.B. 09-04-2010)—also called *Décret « inscriptions »* or *Décret « Simonet »*. In the face of criticism from some Belgian economists (e.g., E. Cantillon, 2013; Maniquet, 2009), the new decree eliminated randomisation in favour of a complicated system of parental preferences managed by a centralised agency, but it retained a quota of 20.4% school places to be allocated to disadvantaged students. It, too, was met with some discontent: since it relied on statistical approximation to calculate priorities, it created problems for a minority of families. Overall, however, it enjoyed a better reputation and had a better political fate than its predecessors (Ryelandt, 2013).

The extent to which the policy was enforced is less clear. One interviewee noted that many in the cabinet to the MoE Simonet were school principals who strongly opposed their Minister's decree. It 'was really a paradoxical situation in which she knew that the rules were not followed' but could not take any action other than 'saying to the principals "well, you should do what the *décret* says…" [laughs] that's Belgium' (B7). Another expert said that having a quota 'does not mean that this 20%, you should put them all in the same classroom. Some schools do just this, there is no *mixité*…' (B6).

Impact

The *Décret « Simonet »* featured in the PISA 2013 international report as a possible policy option to reduce segregation (OECD, 2013g, p. 133). So far, however, there is little evidence that this outcome was achieved. Three reports of the Inter-network Enrolment Committee and an interim report of the Steering Committee found a small 'quivering' of school population

statistics towards greater diversity, but this was limited to certain geographical areas and a few percentage points (Commission de Pilotage, 2013, p. 44; also Ryelandt, 2013). The main impression, which was restated in the final report (Commission de Pilotage, 2014) was one of general stability in equity and achievement since 2007. Two interviewees (B5, B6) suggested that there can be no improvement at the school level so long as there is socio-urban segregation, which is an issue particularly in Brussels.

Policy alignment of Belgian policies to the EPA

There is extensive evidence that the OECD goals and values were acknowledged in many education policies in the CFB, particularly after PISA. As often restated in this research, however, sharing goals (acknowledgement) is no guarantee that specific policy levers will follow OECD recommendations (alignment). Sometimes, alignment may even precede acknowledgement. For instance, policies on teacher training and CPD in the CFB were introduced before a narrative was constructed linking teacher quality with student outcomes (Cattonar, 2010).

Policy alignment in the CFB is summarised in the table below. As in the case of Ireland (Table 5.2), the first column displays the year from the most recent overarching education framework before PISA, the *Décret « missions »* of 1997, until the last PISA cycle at the time of this writing. The T1–P3 columns are the EPA areas (footnote 21, p. 209), and the last column lists the Belgian policies under consideration by year of coming into force, when known, or publication.

Table 6.3: policy convergence towards the EPA in the CFB

Year	T1	T2	E1	E2	E3	E4	E5	Q1	Q2	Q3	Q4	Q5	P1	P2	P3	Policies
1997													1			*Décret « missions »*
1998									2							D. on (pre-)primary education, *D. « discriminations positives »*
1999																
2000													3			D. on teacher training, pedagogical cycles (1)
2001																Core Skills, additional science instruction
2002																Steering Committee, ONE, IFC and D. on CPD, *ouvrir mon quotidien* (primary)
2003																
2004												4				D. on special education, D. addressing school dropout, exclusion and violence
2005																School Contract
2006														5		CEB, external assessments, *ouvrir mon quotidien* (secondary)
2007					6					7						Inspectorate, *Décret « inscriptions »*, D. on school leaders, pedagogical cycles (2), ETNIC, D. reorganising *1er degré*, personalised learning plan
2008																*Décret « mixité sociale »*, pedagogical cycles (3)
2009																
2010																*Décret « Simonet »*
2011																
2012																

Colour coding:

Contrary	No policy	Acknowledgement	Alignment / Enactment	Implementation

The cell colours represent policy alignment: the white colour means that there is no national policy clearly attempting to intervene on that EPA category; red is for policies that are contrary to OECD recommendations; white means that there is no national policy clearly attempting to intervene on that policy area; yellow stands for interventions purporting or aspiring to be in line with the OECD goals; blue is for policies actually aligned with the EPA, whether intentionally or not and regardless of whether they managed to be implemented effectively; and green represents policies that are not only in line with OECD recommendations on paper, but for which there is plausible evidence of implementation.

Colour shades signal when policies first addressed the EPA (dark shade) and how long for (light shade). Once again, this does not take into account whether implementation was effective but only the fact that, at a certain point in time, a policy agreeing (dark blue/green) or disagreeing (dark red) with the EPA came into force, and that policy was "active" for a number of years afterwards (light blue, green or red). Notice that yellow comes in one shade only because it serves simply to flag up purported but not actual alignment.

Finally, the numbered cells index some policies that are followed up in the following section to make more general comments.

Trends in policy convergence towards OECD recommendations

Most cells in Table 6.3 are blue, suggesting that the CFB and the OECD generally agree on educational matters, especially after the School Contract of 2005. Compared with Ireland, it was more difficult to gather evidence that policies were implemented at the classroom level. This is because the fragmentation of the education system and the high autonomy enjoyed by POs made large-scale evaluations more the exception than the rule. As Mangez & Cattonar (2009) explain, Belgian society traditionally relied on élites (for instance, religious and secular powers) achieving results by negotiating, compromising and generally 'not meddling in the affairs [… of] one's neighbours' (p. 18). In this context, systemic or 'objective' knowledge was of little political utility. Quantitative evaluations had no place in public policy before PISA.

This created two hurdles in this research: determining whether a policy was implemented, and determining whether it was implemented as intended (plus, of course, deciding what impact it had if implemented). In the first case, a decision was reached after considering who was in charge of the implementation (schools, *réseaux* or central government): the more centralised the procedure, the more likely the implementation. It was harder to decide whether a policy retained the spirit of the legislator—an issue referred to by an expert as the difference between achieving 'policy *impact on*' versus 'policy *integration into*' classroom practice (B2). This depends on the strength of the coupling between different levels of the education system

(Scheerens et al., 2015), which in the case of the CFB is very loose. In practice, this meant that many cells in the table could have been shaded in blue or green according to one's interpretation of the evidence.

The numbers in the cells of Table 6.3 exemplify some of the problems with the classification of Belgian policies. The first concerns the supplementary year introduced by the *Décret « missions »*. Its purpose was to provide a modified curriculum for underachieving students until they achieved the Core Skills (i.e., they were awarded the CEB), so at least in its intentions it was a supporting mechanism. It also grouped underachieving students forcing them to spend an extra year in education, and in this respect was a form of grade retention. Consider, however, that grade retention was already practised, so if the supplementary year had been integrated into school practice as a supporting—rather than a sanctioning—mechanism, it could have been considered an improvement compared to the previous situation.

The final decision to assign a blue colour to the cell (alignment, but not implementation) was reached after evaluating the decrees on positive discrimination (point 2 in the table). As a centralised intervention, the likelihood that resources were indeed allocated more equitably was higher and therefore, when in 2004 a SES-based formula was introduced to give extra funding to schools in need, the cell turned to green—though it could still be argued that one cannot know whether funds were used for equity purposes.

Following these lines of reasoning, pedagogical cycles (3) were intended as a way to support low achievers (column P1) and simultaneously address grade repetition (column E3), but there is little evidence of integration and therefore the cell was shaded in blue. Similarly, there is currently an 'overload of measures and devices' to tackle school dropout, exclusion and violence, which creates a 'parcelling out of the actions to take' (ASPE, 2013, p. 17) and makes it difficult to gauge whether these policies (4) have been implemented as expected on a large scale.

Number 5 refers to the examination leading to the CEB. As argued earlier in this chapter, its status as a centralised external exit examination is disputable, and this is the reason behind the yellow label: there is acknowledgement of the goals and values of the OECD, but hardly any alignment with the EPA. Nevertheless, in combination with decrees affecting the organisation of the *1er degré* (e.g., D. 07-12-2007, M.B. 26-02-2008) it is thought to have addressed some inequalities by reducing the amount of low achievers, the gender gap and the native/non-native gap in reading (hence the green shading at point 6, see Lafontaine, 2014).

The reason why the cells in column E3 remain light red-shaded after 2007 is because the combined effects of the CEB regulations and the decrees were only 'small steps in the right direction' (Lafontaine, 2014): previously, pupils failing to acquire the CEB were streamed

into pre-vocational tracks without any support; after the reforms of 2006–2007, these pupils are streamed into the *1er degré différencié*, which has a closer curriculum to the *1er degré commun* and makes some pedagogical provisions. However, horizontal and vertical segregation mechanisms (i.e., curriculum differentiation and grade retention) are still in place.

Finally, the transition from blue to green in column Q3 in 2007 (7) signals the consolidation of accountability policies in the CFB. Surely, school autonomy in resource allocation (including teacher pay) was still limited and public posting of achievement data was forbidden. On the other hand, autonomy in curriculum and pedagogy was already high and external evaluations were introduced through examinations, assessments, school education plans and inspections. Even if these measures did not have a large impact on teacher practice or student outcomes, they represented a considerable disturbance of the system.

The impact of EPA-aligned policies on achievement and equity

Of the many policies theoretically in line with OECD recommendations, only one has been thoroughly and effectively implemented: the change in enrolment procedures in the *1er degré* (the decrees *« mixité sociale »* and *« Simonet »*). Thus far, the reform has produced little measurable improvement but there is also some evidence of greater diversity within schools.

According to the OECD argument, PISA should be able to detect changes in the performance and equity levels of the CFB as a result of its policies. This argument was put to the test using Willms (2006) LSA-based policy evaluation framework. Recall that the framework states that the effects of 'universal' interventions should be captured by changes in mean country outcomes, 'SES-targeted' interventions by the Strength and Slope of the socio-economic gradient, and that 'performance-targeted' interventions should lower the percentage of students scoring below proficiency levels 1 or 2 (Table 6.4).

Table 6.4: a framework to detect the effect of policies in the CFB on PISA outcomes

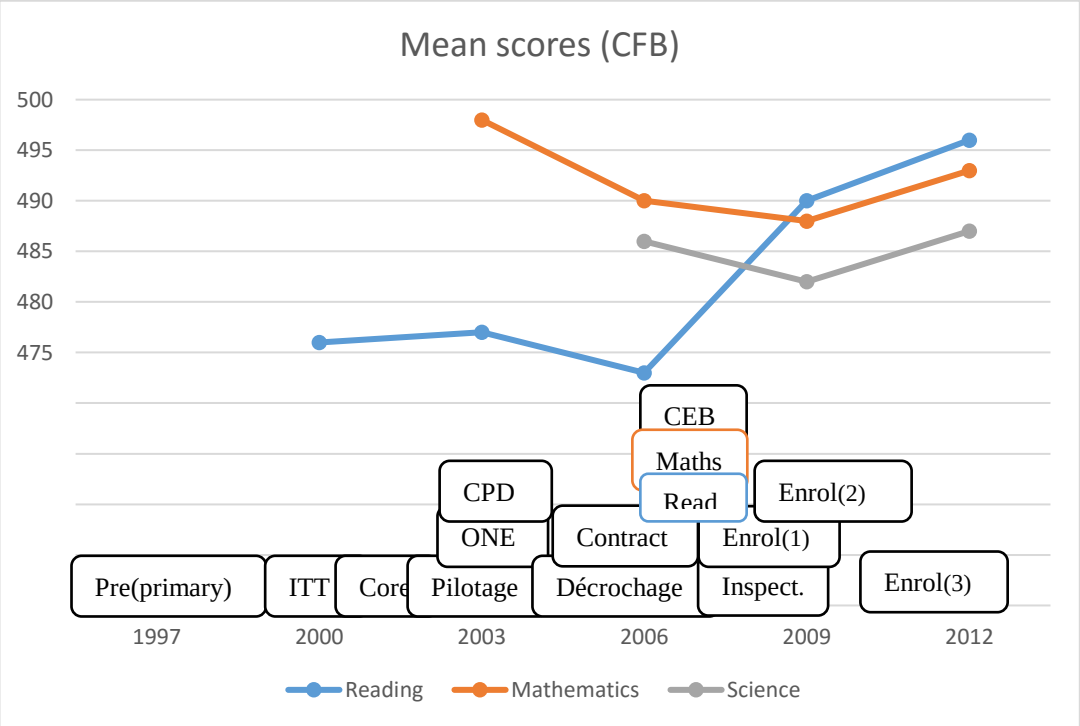
Type and variable	Sample policies in the CFB	Abbreviation
Universal interventions *strive to increase the educational performance of <u>all children</u> through reforms that are <u>applied equally across the schooling system</u>.* ***Variable:*** *mean country performance*	D. organising (pre-)primary education	(Pre)primary
	Decree on teacher training	ITT
	Core Skills	Core
	Additional reading / mathematics instruction	Read / Maths
	Steering Committee	Pilotage
	Office de la Naissance et de l'Enfance	ONE
	Institut Formation en cours de Carrière	CPD
	D. on dropout, exclusion and violence	Décrochage
	School Contract	Contract
	CEB	CEB
	Reform of the Inspectorate	Inspect.
	Décret « inscriptions »	Enrol(1)
	Décret « mixité sociale »	Enrol(2)
	Décret « Simonet »	Enrol(3)
	D. on school leaders	Lead.

SES-targeted interventions *aim to improve the educational performance of students with low socioeconomic status by providing a specialised curriculum or additional instructional resources.* ***Variable:*** *Strength and Slope of the socio-economic gradient*	Positive discriminations (1,2,3) School Contract Décret « mixité sociale » Décret « Simonet »	D+(1)/(2)/(3) Contract Enrol(2) Enrol(3)
Performance-targeted interventions *provide a specialised curriculum or additional instructional resources for particular students based on their levels of academic performance.* ***Variable:*** *Percentage of pupils below level 2*	Pedagogical cycles (1,2,3) School Contract Personalised learning plan for struggling students Decrees on 1er degré	PC(1)/(2)/(3) Contract PIA 1er

Source: adapted from OECD (2013e, pp. 108–109); Willms (2006)

Figure 6.1 is a timeline of universal interventions and country mean scores since 1995. Interventions focusing on raising performance in a specific domain were colour-coded.

Figure 6.1: Universal interventions and performance in the CFB



Unlike the case of Ireland, country mean trends in the CFB converge. While Mathematics and Science have been rather stable, Reading performance has increased by about 21 points, or a fifth of a PISA standard deviation. The ASPE has argued that the improvement, which affected the lowest performers in particular, could be due to three factors: an additional hour of French instruction introduced in 2006; the students' greater familiarity with PISA-type questions thanks to the external assessments; and the effect of the *année différenciée*, which

forces even the weakest students (those who did not attain the CEB at the end of primary education) to reach basic skills in French and Mathematics (Lafontaine & Baye, 2012; Lafontaine, 2014).

This possibility can be critiqued. Firstly, the three factors should also have influenced the Mathematics domain (mathematics benefitted from additional instruction time, too). Secondly, after the introduction of the *année différenciée*, there were more students in the *1er degré supplémentaire* and they performed better. Perhaps, higher performance does not capture the effects of instruction but the increase in number of students who had the ability to perform at level 3 or 4 in PISA and were retained because they did not attain the CEB. Finally, Reading scores aligned with Mathematics and Science scores as their standard error shrunk to almost half of its magnitude in 2000 (data not shown): PISA has become more accurate over time, so earlier Reading scores may have been underestimated.

Consider that the confidence interval at the 5% level for Reading 2000 was 462–490 and the mean score in 2012 was 497. It can be shown that if the 2012 value had been one point lower, the improvement between 2000 and 2012 would have been non-significant at the 5% level. Once again, this is not to deny that a change happened, but given the many issues in PISA reliability—especially in Reading—highlighted in Chapter 4, a certain caution when interpreting these results is recommended.

Figure 6.2 and Figure 6.3 show the Slope and the Strength of the socio-economic gradient in the CFB.

Figure 6.2: SES-targeted interventions and the Slope of the socio-economic gradient in the CFB

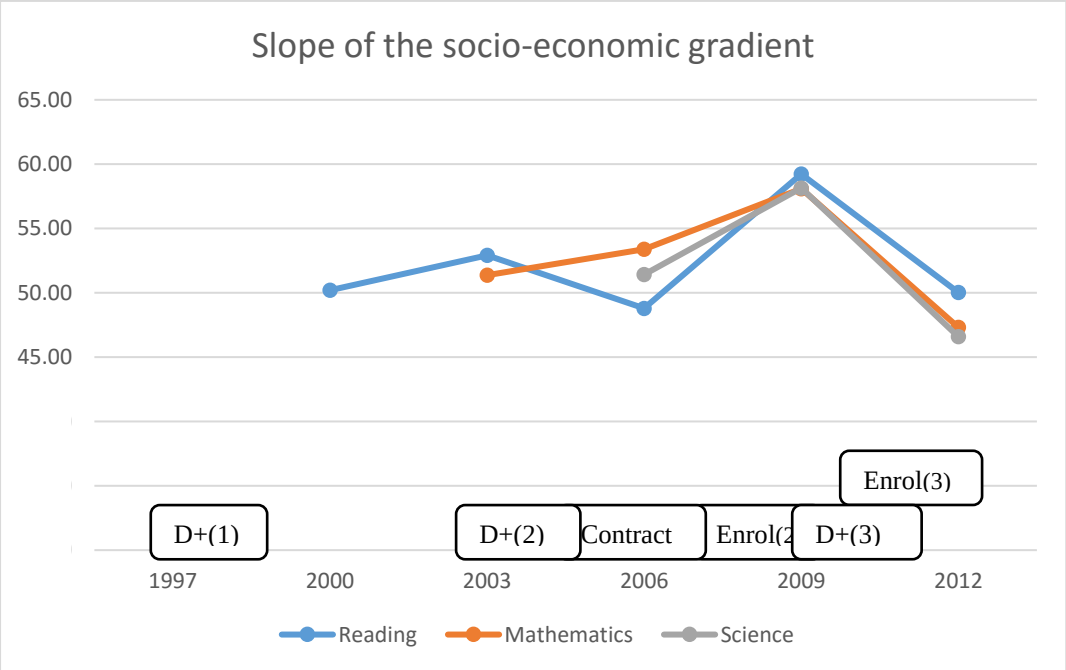
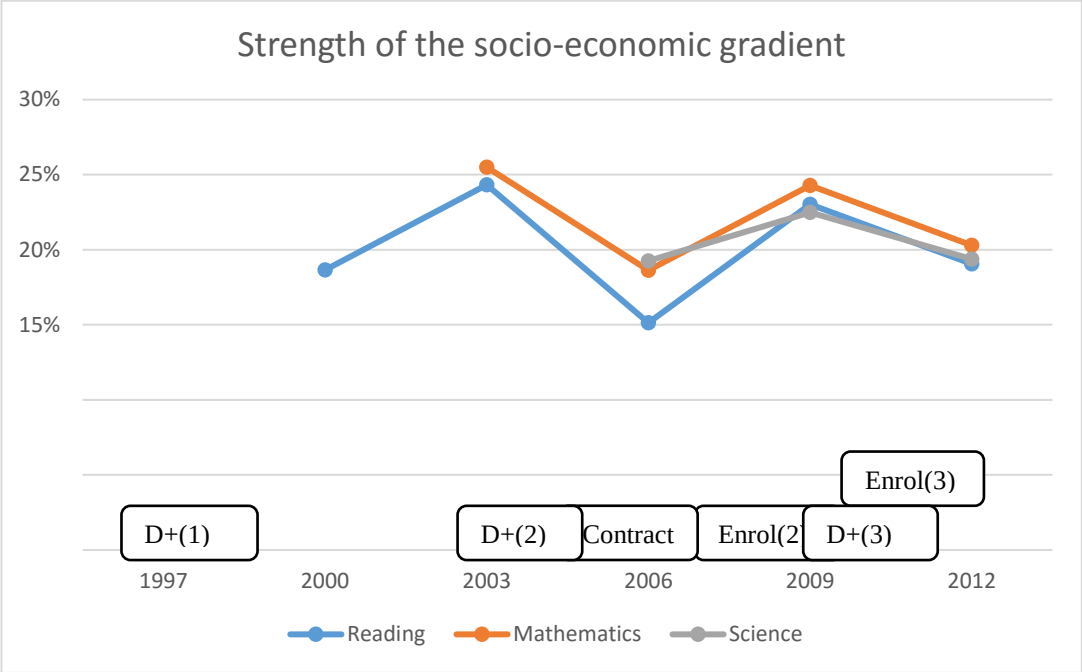


Figure 6.3: SES-targeted interventions and the Strength of the socio-economic gradient in the CFB



Source: Author's calculations using the PISA database. 2003 data for the CFB were not available in the OECD database and had to be extrapolated, which led to missing data. Since 2006, the CFB and the German-speaking Community data were merged.

A consistent finding emerging from documents and interviews was that inequality levels did not change in the last 15 years; if anything, retention levels have worsened (Baye et al., 2014). Data from PISA agree with information gathered on site: a difference of one standard deviation in socio-economic status corresponds to an average score difference of 52 points, and this relationship captures about 20% of the variance (higher than the international average 2000—2012, which is 13%).

Figure 6.4 and Table 6.5 link performance-targeted interventions to the performances of the highest- and lowest-achieving students.

Figure 6.4: Performance-targeted interventions and performance in the CFB

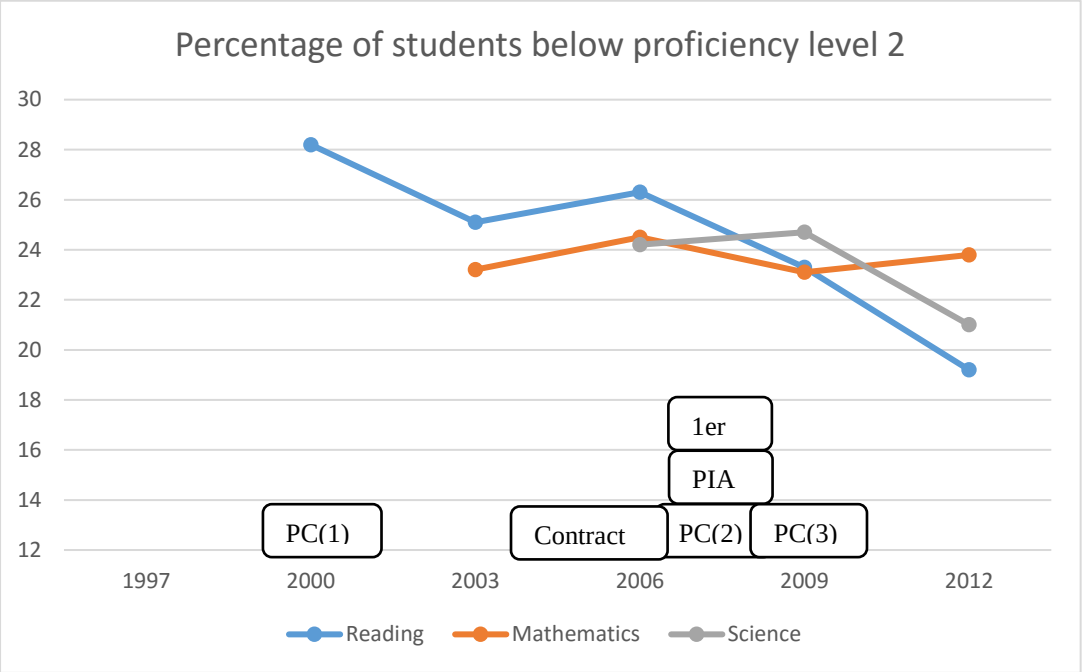


Table 6.5: Changes in the performance of the highest and lowest achievers in the CFB between the first comparable assessment and PISA 2012

	Scores	**% < lv2**	**% > lv 4**	**10th score**	**90th score**
R (2012-2000)	+21	-9	NS	+36	NS
M (2012-2003)	NS	NS	-4.2	NS	-17
S (2012-2006)	NS	NS	-2.1	NS	NS

Source: PISA database and national reports for the CFB. Standard errors for PISA 2000 were not available, therefore the average of following cycles was imputed. "Scores" = change in country mean scores; "% < lv2" and "% > lv4" = change in the percentage of students performing below the performance level 2 or above level 4; "10th score" and "90th score" = change in the scores of the 10th or the 90th student percentiles; "NS" = Non-significant at the 5% level)

Figure 6.4 shows that there was a constant decrease in the percentage of students below level 2 in Reading and Science, but not in Mathematics. Table 6.5 adds that these changes were accompanied by performance improvements only in the case of Reading, whereas there was a decrease in the performance of the higher achievers in Mathematics and Science. It was argued above that improvement in reading may be due to stronger performers being retained in lower secondary education. Admittedly, this explanation conflicts with the -17 difference observed at the 90th percentile of PISA Mathematics: grade retention would explain it, but it is unlikely that retained students were top achievers.

Belgian researchers claimed that reading comprehension was '*the* big project in the years 2000–2010' (Baye et al., 2010, p. 17, emphasis in the text), but there was little evidence available from this research that the project had been implemented coherently and successfully on a national scale. Besides, various interventions affecting the vertical and horizontal

streaming of students have altered the distribution of the PISA population among grades and tracks.

PIRLS 2011, which also focused on reading, showed no changes in performance at the primary level and a situation which was 'way more worrying' than that at the secondary level, with a large proportion of low achievers (Schillings, Hindryckx, Dupont, Matoul, & Lafontaine, 2012, p. 7). If the project to improve reading had been successful, it is unclear how the PISA and PIRLS cohorts may have experienced such a different instruction that would account for their different results. This gives some credit to the explanation reported above that improvement in reading was not driven by subject-specific reforms, but by the reforms of lower secondary education of 2006–2007, which would affect the PISA but not the PIRLS cohort (Lafontaine, 2014; and personal communication of 9 February 2016).

Whether this is the case will appear more clearly after the next PISA cycle, but it should be stressed that, once again, Willms' framework did not make it any easier to uncover PISA-policy relationships. In fact, the evaluation in this chapter is evidence of the difficulties of linking interventions and consequences when policies are examined in detail. And yet, the OECD and some econometric research still recur to simplifications such as "number of policies in one domain" or "number of SES-targeted interventions" for policy evaluation as if it were possible to draw valid conclusions from these variables.

Conclusions to Chapter 6

Evidence of country alignment with OECD recommendations

In the online questionnaire, the national expert agreed with Hopkins, Pennock, & Ritzen (2008) that policy alignment in the CFB was 'medium'. If one considers EPA reception only, this is certainly an understatement. PISA results were highly mediatised and known by all stakeholders. Consistent with the theory, there is extensive evidence that policymakers extrapolated results to support their arguments. Some referred to this practice as 'reading recommendations according to an ideological prism' (B5), some were more openly critical: 'And so PISA becomes like a potluck: people find what they bring' (Romainville, 2002, p. 1).

Two parallel narratives appear to have developed in the CFB, and both use the PISA findings that best support their claims: on the one hand, players broadly linked to the political left focus on equity and the high proportion of low achieving students to request structural changes; on the other hand, the Catholic network and the political right focus on performance and the low proportion of high achieving students to request more autonomy. Even their reference societies differ: Scandinavian countries and Poland for the former, the Netherlands and the Flanders for the latter (E. Mangez & Cattonar, 2009; also B3 and B4).

Because of its active role in policymaking, the ASPE was often accused by its direct competitors of being politically motivated (Cattonar et al., 2009)—unlike the OECD, which is generally considered *super partes* and whose data is seen as reputable and objective. This is interesting, as it was shown in this chapter that there is usually a remarkable alignment between OECD and ASPE reports, meaning that they should equally be slated of partiality or praised for their objectivity. From its part, the ASPE is aware of its responsibility as a policy advisor but also, together with some informants (e.g., B3), regards PISA as a diagnostic tool for the use of researchers whose findings do not necessarily entail policy decisions.

In terms of acknowledgement, PISA entered the political discourse representing international standards to which Belgian policies have increasingly aspired to conform. Acknowledgement of the OECD goals and values was expressed mainly through policies addressing equity, educational governance and student and system evaluation. This was a novelty for the CFB (B4). A former member of the Steering Committee said: 'the first time I talked about an external examination, the common CEB, they looked at me [with an expression] saying "where are you from??" […] and then we did it. When we came with the idea of the CE1D at the end of the *2e degré*… I've never had it as easy to negotiate a text since everyone agreed. Everyone said "well, of course [we should do it]!"' (B-).

It was shown earlier that there is considerable alignment between Belgian policies and the EPA, particularly since the School Contract of 2005. There have been also attempts to address equity issues, although in many instances equitable aims were counterbalanced by segregating outcomes. Future policies have the potential to be even more strongly informed by the EPA: for instance, the new "Pact for an outstanding education" (http://www.pactedexcellence.be, accessed 16 April 2016) is being developed in collaboration with McKinsey Belgium and the OECD (B6, B7).

As was the case in Ireland, one of the most-debated but least-affected areas was the teacher continuum. The decrees in the early 2000s 'were not major reforms', and there were few policies 'trying to influence what teachers do in the classroom' (B3). The problem with reforms of teaching, it was argued (e.g., B6), is that they would affect the existing status quo among teachers, teacher training providers and *réseaux*: in other words, any proposed intervention would be resisted by at least half of the stakeholders.

Evidence of EPA-aligned policy effects

The ASPE has claimed that Reading performance has increased thanks to the combined effects of examination policies (the CEB), support mechanisms for low performers (an extra hour of French and the *année différentiée*) and a greater familiarity with PISA. Of these factors, the only one that can be said to be somewhat in line with the EPA is the *année différentiée*, which

"ticks the box" as a support mechanism but does not change existing streaming arrangements. Moreover, the lack of increase in Mathematics and the standard error on Belgian performance suggest caution in accepting the argument of the ASPE.

On matters of equity, despite years of consultations and several attempts, EPA-aligned policies addressing systemic inequalities have produced little improvement thus far, whereas grade repetition appears to be worsening. Some mechanisms are deeply rooted in the system. Most teachers and parents support grade repetition as a useful measure to address behavioural and academic problems (Dupriez & Dumay, 2008, also B6). Cornet & Dupriez (2004) found that some teachers reject the idea that social inequalities could be visible from a young age, and when presented with the evidence they failed to link national trends to the reality of their own classrooms. This creates a 'culture of failure' in the CFB whereby 'normal' students proceed at an expected pace and everyone else is retained (Baye et al., 2014).

Some opt-in projects were reportedly more successful than systemic policies in addressing these issues. An example is *Décolâge*, which attempts to reduce retention in primary education by providing teachers practical and cultural alternatives to it (e.g., B5, B6, B7). The issue then becomes upscaling ('we are world leaders in pilot projects', B6). Besides, measures to change the culture of grade repetition would not be sufficient when retaining a child is instrumental to increase the school population and access more funds, as it was claimed to be the case in some kindergartens (B6).

If aligning to the EPA has not been a "game changer" in the CFB with respect to performance of equity, there was at least some consensus among interviewees that PISA was a useful instrument for policymaking, especially as a diagnostic tool: 'it taught us a lot' (B2). The fact that findings can be used 'to support A as well as the opposite of A' (Figazzolo, 2009, p. 28) was considered more an indication of its quality than a limitation. PISA was likened to a thermometer (B2, B4, B6) producing 'solid empirical data' (B3), whereas national tests 'can always be suspected of being subjective' (B6). To some, PISA created 'a cult of international evaluations' because it was the only objective instrument available (B5), and in general there were so many stakeholders using PISA for their purposes that criticism 'remained peripheral' (B3).

Limitations and future directions

All research activities have to compromise between scope, scale and accuracy, and this study was no exception. It was the first time an attempt to reconstruct the whole OECD policy advice and link it with national policies and country outcomes was made. New and valuable data were gathered, but not all parts of this study were equally effective. In this section, some limitations are addressed and suggestions for further studies are provided.

The main limitation when reconstructing the EPA (RQ1) was that there is no unique source from which it originates and no unique audience. Over 100 OECD publications appear each year; most of them have multiple authors, contributors and editors and are directed not just at policymakers, but at researchers and other stakeholders. Online databases have no recognisable author, and even the PISA reports, which are produced by the Secretariat, are a collective enterprise and are reviewed by the PGB before publication. As was argued in Chapter 1, there is no real distinction between an international and a national level when it comes to the relationship between the OECD and individual member countries. This is because of the very nature of the organisation: the OECD *is* its member states but it is also more than the sum of its parts; national experts may sit in committees taking decisions which affect the international community; the Secretariat has some autonomy but not full autonomy.

As a result, whenever a recommendation was analysed it was often unclear who was talking and to whom. Was a piece of advice being pushed by a research agenda or policy agenda? Was it meant to be heard by researchers, teachers or by policymakers? Knowing this could have helped to focus the evaluation on the most pressing issues. Instead, during the study all recommendations felt relevant, all were given the same weight even though some might have been less prominent in the OECD discourses of the last fifteen years.

This limitation was partly mitigated by adopting the OECD categorisations of policy options to identify key areas and thanks to the validation of the EPA reconstruction (see Table 3.2) by Andreas Schleicher. Nevertheless, a different clustering of policy recommendations or a different way to approach the synthesis (for instance by interviewing more experts within the OECD to reconstruct the EPA) would have altered the EPA presented in this thesis and perhaps some of the findings on policy alignment (RQ2).

On this matter, both the email survey and the online questionnaire (the PAIQ) had limitations. To increase the response rate, the email survey asked only one question, therefore it could not capture trends over time. It is likely that each respondent interpreted the question slightly differently. This is common with expert judgements, and there was some evidence suggesting that experts struggled to decide whether policy alignment should take

implementation into consideration. Moreover, it is not known how much effort each put into responding, and answers tended to gravitate towards the central option.

Ideally, one would have wanted to ask more questions to more experts to be able to rate their consistency, and to follow them up to probe effort and understanding. It was once again a matter of trading off depth for breadth, but a follow-up attempt was made through the PAIQ. The strength of the PAIQ was that it was straightforward to complete and it returned simple but thorough information. The response rate was also relatively good (about 40% of the email respondents completed it). However, the accuracy of the PAIQ relied on the accuracy of the EPA conceptualisation and, as stated above, this latter was difficult to evaluate. The four areas and 15 subcategories in which policy options were grouped into were informed by OECD publications but still depended on this author's judgement.

The fuzzy boundaries of the EPA and its scarce internal coherence had repercussions also on this author's ability to evaluate policy alignment in the two case studies. The concept of "delta-convergence" was appealing, but the model promoted by the OECD was not suited for a quantitative approach. A solution was found by visualising policy convergence on a time map (see Table 5.2 for Ireland and Table 6.3 for the CFB), but even then it difficult to assign some interventions to a definite category since they agreed with one recommendation and went against another. At least, this confirms that this author's scepticism towards the additive approach of Braga et al. (2013) for quantifying policy impact was legitimate.

An analysis of the role of national project managers (NPMs) and other policy brokers was outside the scope of this research, but it was clear from the two study visits that the ERC and the ASPE did more than reproduce OECD recommendations within local boundaries. This was not a huge limitation, because changes in the EPA due to the NPMs' contribution were taken into account. Nevertheless, future developments of this research could look into this issue, drawing from existing literature on knowledge brokers (Meyer, 2010; Sriprakash & Mukhopadhyay, 2015), on the political role of "independent" policy advisors (Craft & Howlett, 2013; Van Damme, Bossens, Brans, & Fobé, 2014), and on the relationship between the ability of countries to make sense of international policy advice and policy convergence (Howlett & Joshi-Koop, 2011).

Finally, it was hard to locate or produce solid evidence of policy effectiveness (RQ3), as is often the case in policy evaluation. In Chapter 4, the correlational and multi-level analyses returned some valuable information, but expert responses were too few to provide conclusive findings on the influence of EPA-alignment on achievement or equity. The association between the EPA Index and PISA rankings was probably driven by an interacting variable, and while small effect sizes and lack of data were expected, they were still somewhat

disappointing. Although the information from the NPMs was valuable and provided contextual information, there have been too few PISA cycles to observe clear trends.

Willms' (2006) framework was adopted as an additional theoretical instrument to facilitate policy evaluation because it is used by the OECD, but in the end it did not make policy effects much clearer. Country policies seemed to be always unrelated to country trends when they were grouped according to the framework, with the exception perhaps of the impact of Belgian policies on the *1er degré*, which may have been captured by the decrease in percentage of low achievers as expected by the framework (Figure 6.4). Notice that this disconnect was present also when independence evidence of policy effectiveness existed. For instance, the DEIS policy in Ireland improved equity, but this was not captured by changes in the PISA socio-economic gradient as posited by the OECD and Willms' framework.

If one considers the positive correlations observed in Chapter 4 between changes in pre-primary enrolment and changes in PISA outcomes as an indication of a real underlying policy effect, then the question becomes what types of policy is PISA more sensitive to, and whether Willms' framework is the best way to categorise policy outputs.

Despite these limitations, this research adds to the current body of knowledge through its attempt to link the production of educational knowledge (the EPA) to consequences and outcomes, while providing the theoretical and methodological instruments to extend this work to other countries and future assessments. It also addressed two requests expressed by other researchers and outlined in Chapter 1.

Baird et al. (2011) proposed that 'a more thorough study of the temporal relationship between policy thrusts and PISA results [… could] shed light on the plausibility of the claim that PISA causes policy' (p. 2). Evidence from Ireland and the CFB suggests that this possibility should not be excluded, especially in recent years. Specifically, PISA results might have directed the attention to policy advice that was just waiting to be noticed and integrated into national plans.

Hanberger (2014) stated that 'whether and how PISA has helped develop education (systems) and its consequences for school practice merit further research' (p. 176). PISA has certainly helped to develop a culture of evaluation and accountability, but both Ireland and the CFB reinterpreted this culture to suit their existing practice. Since neither education system had a tradition of large-scale evaluation and accountability, the outcome was a top-down approach whereby policies reproducing the narrative of the OECD were introduced but had to face national contexts in their application. School inspections became more frequent and teachers felt under attack, however no real sanctioning or formative measures followed from them. Consequences for school practice may have included an increased administrative burden

or a sense of vulnerability on the teachers' part, but at the time of writing the impact of PISA in terms of evaluation and accountability was more in words than in practice.

Future research might want to update the EPA and the multilevel analysis in light of new data whilst keeping the current methodology, and focus its attention on other countries. Because policy intentions often do not transpire from published documents, the use of educational experts is still encouraged. Where possible, however, expert opinions should be collected from different time points (either longitudinally during the research or through historical analysis) to account for policy changes. Expert opinions sometimes change when the policies do. In this study, time was factored in when looking at policies and policy advice, but it was given somewhat less prominence when reporting informed opinions (for instance, semi-structured interviews did not include contrasting questions such as "What did you think at the time? What do you think now?").

Conclusions

This research sought to evaluate the OECD education policy advice (EPA) by assessing the extent to which following OECD policy recommendations can help countries to achieve the goals of greater performance and equity in education. The evaluation addressed the "theory and practice" of the OECD policy advice: on the one hand, what the OECD says countries should do and whether this is reasonable advice; on the other hand, which recommendations find actual application on a national level and whether this can be linked to improvements in quality and equity.

The rationale for the evaluation was that the OECD has gained prominence as a policy advisor in educational matters, particularly after PISA. PISA has become a well-known "brand" among policymakers, and the education reforms of participating countries have more or less implicitly purported to align with the OECD goals or have aimed to rectify country underperformance in the assessment. This has led to a strand of research investigating the use, usefulness and effects of PISA: how different countries prepare and respond to the survey; whether it is a valid and reliable measure of student abilities; and how international surveys such as PISA promote a culture of accountability at the country level.

By placing the spotlight on PISA, this literature has analysed only the more superficial and "neo-liberal" aspects of the EPA, such as advice to deregulate and decentralise. National policies thought to be in line with OECD recommendations were considered as examples of the globalising effects of neo-liberalism on education systems through PISA. What was missing in the literature was a systematic and comprehensive analysis of what the OECD advises that countries should do to improve education, beside the "neo-liberal" suggestions. It was unclear whether policies thought to be in line with the EPA were truly so or simply tried to adhere to some vague international standards, or whether EPA-aligned policies have helped countries to increase the performance and equity of their education systems as envisaged by the OECD.

If the EPA is advice worth following, it should be possible to detect improvements on a national or international level when it is followed. Therefore, this research attempted to fill the gaps in the literature by assessing the quality of the EPA and looking for connections between EPA-informed policies, country performance in the three PISA literacies of Reading, Mathematics and Science, and country equity in education.

The evaluation of the education policy advice was guided by four questions:

RQ1) Is the OECD education policy advice "sound" advice?

RQ2) To what extent do countries follow the OECD education policy advice in their national policies?

RQ3) What is the evidence that education policies informed by OECD recommendations influenced country performance?

RQ4) Considering the evidence, how valuable is the OECD education policy advice?

RQ1 aimed to reconstruct the EPA, analyse whether it was coherent, consistent and supported by evidence (including PISA evidence). This was the "theoretical" aspect of the OECD policy advice mentioned above. Its "practical" side was investigated in RQ2 and RQ3. RQ2 was about the use of the EPA. Acknowledgement of OECD goals by policymakers was separated from policy alignment with the EPA, as only interventions actually in line with the EPA could provide evidence that following it made a difference to performance and equity. RQ3 searched for this evidence by taking policy implementation into consideration. Finally, the last research question was used to summarise the evaluation by looking at the overall value of the EPA. Here, "value" means that the EPA is advice worth following either because there is evidence that it *could* work (RQ1) or because there is evidence that it *did* work (RQ2 and 3).

In the continuation of this chapter, findings are first broken down by research question, and then compared against the research hypothesis laid out in Chapter 2.

RQ1: Is the OECD education policy advice "sound" advice?

After analysing OECD publications and receiving feedback from the Director of PISA, Andreas Schleicher, this research identified four areas in which the OECD provides policy advice: educational time, equity, quality of provision and student performance. Across the four areas, a total of 15 different policy recommendations were given, spanning from increasing instruction time, to reducing school competition, to promoting cooperation between schools and parents (Table 3.2).

The policy options provided by the OECD are rather specific considering the scale of the endeavour, but the framework itself is not very cohesive. Even though the analysis concerned *education* policy advice, it was hard to identify clear boundaries between recommendations targeting education policymakers and those could may be better directed to policymakers in other areas of country governance (such as welfare, finance, or infrastructures).

One implication of this can be exemplified by looking at the whole school evaluation strategy in Ireland. The Inspectorate claimed that the strategy had been informed by the OECD education advice, but a closer analysis revealed that the it had started to take shape before PISA and had been developed independently. It did borrow an overarching narrative of accountability and efficiency from the OECD, but when it did, it borrowed recommendations directed at policymakers outside education. The loan was possible because it took concepts

such as “accountability” and “efficiency”, which are not specific to one policy domain and are therefore highly versatile multi-vocal symbols.

As a Belgian policymaker put it: ‘in a more or less conscious fashion, the OECD discourse comes in the ideas that are in the air at the time, […] we know that there are some ideas that come and that everybody says at a certain time without knowing who said it among them’ (B5). The risk with this is that, when making policy, vague concepts may be preferred over potentially useful domain-specific policy recommendations.

The multi-vocal nature of the OECD’s educational concepts and policy options means that the extent to which the EPA can act as an organic whole, rather than as a collection of stand-alone suggestions, depends on the reader. The framework is a ‘patchwork of loosely connected topics [… that] have been strung together to provide a minimum semblance of coherence’ (Henry et al., 2001, p. 52). Any further systematisation is added by the ideological position and prior knowledge of policymakers. To paraphrase another Belgian expert (Romainville, 2002), the OECD policy advice is like a potluck, in which you find what you bring.

Another implication of the (scarce) internal coherence of the EPA is that OECD recommendations in educational matters are not simply a reproduction of a “neo-liberal” narrative. The double quotation marks are necessary because there is no single accepted definition of “neo-liberalism”. For some it is ‘synonymous with unfettered capitalism and an economic imperialism that aims to subject all spheres of society to the logic of the markets. Others suspect the term to be useless except for polemical purposes’ (Biebricher, 2015, p. 256). Neo-liberalism involves deregulation, market mechanisms, privatisation, accountability, performativity and efficiency, but there are regional and local variations on the form it can take (Connell & Dados, 2014).

It is evident that the EPA is not fully “neo-liberal” in this sense. Demands for performativity and efficiency are matched by calls for greater equity, which also involve cooperation and reduction of market-based competition. In this respect, the EPA reflects both the strong social democratic positions of some OECD members and the CERI (see Chapter 1), and a recurrent identity crisis of the OECD (Marcussen & Trondal, 2011), which sets the OECD aside from other international organisations such as The World Bank (Mahon, 2010) and is exemplified by the lack of clear objectives for PISA (e.g., Hopkins et al., 2008) other than measuring what can be measured.

It was also shown that the EPA tends to be unaffected by PISA findings. On the one hand, this is because PISA results are stable, but on the other hand it is due to the fact that OECD recommendations are informed by a variety of sources—even when they are presented in PISA reports and supposedly follow from PISA evidence. Furthermore, the OECD has often made instrumental use of PISA just as countries do. Data from PISA are regularly presented and

interpreted to confirm, rather than challenge, pre-existing policy stances. Nevertheless, in recent years the OECD has substantially increased the grounding of PISA in research, which might progressively affect the content of its recommendations.

This research also investigated the soundness of the EPA. In formal logic, an argument is sound if the conclusion follows from the premises and the premises are true. However, it was argued in Chapter 2 that the EPA is a practical argument using appeals to shared goals and values as well as presumptions to tell policymakers where and how to intervene. "Presumptions" are provisional conclusions drawn from some premises when 'there is no sufficient evidence to show that the proposition is false' (Fairclough & Fairclough, 2012, p. 39; quoting Walton, 2006a, p. 72). Therefore, the soundness of the EPA is not so much about its truthfulness or falsehood, but rather the extent to which it is supported by evidence. In other words, what has the OECD been basing its policy advice on? And how does the OECD know that what it suggests will produce any positive effect?

After reviewing selected recommendations from the four main educational areas (educational time, equity, quality of provision and performance), four interrelated issues could be identified. On a few occasions, the EPA was indeed backed by strong evidence from the educational literature but this was not capitalised on by the OECD: the quality and strength of the EPA did not match the quality and strength of the evidence. This was the case with pre-primary education, CPD and streaming practices. It was suggested that this may be due to the prevalence of econometric experts within the organisation, who may pay less attention to topics that traditionally belong to the field of education.

A second issue was the lack of theorisation of some areas of the EPA such as the gender gap. The concept was poorly defined; therefore, the policy advice did not address known issues such as the different mechanisms behind gender gap in secondary versus higher education. This corroborated the impression that despite the aura of objectivity and scientific rationality that the OECD wants to convey, its policy recommendations may or may not be grounded in solid scientific evidence. The OECD appears to have started to address this second issue, meaning that the EPA might become more accurate in the future (provided it also starts drawing from evidence more frequently).

A third issue was when the EPA itself changed over time not because the evidence did, but because the OECD had yet to decide what stance to take on a topic. This was somewhat linked to a poor theorisation of the mechanisms behind certain interventions mentioned above—i.e., what should be done and why it should work. Examples were cited about school vouchers and school autonomy. School vouchers are often unequitable and ineffective, but some OECD publications treated them more favourably than others. This suggests that the OECD does not have one clear position or proposed strategy on vouchers, so its many internal

voices (such as the authors of papers and reports) have enjoyed a certain freedom on how to present them.

School autonomy, in contrast, was treated as a single concept at the beginning of PISA. Although more than one indicator measured different types of autonomy (resources, curriculum, assessment, staffing), the OECD did not explain why it had chosen to measure those aspects or how they interacted except from a purely statistical perspective. At the beginning, the OECD suggested that autonomy in resources allocation could be a viable policy option, but the spotlight later moved on to curricular autonomy. At the same time, a connection was established between autonomy and accountability—perhaps drawing from econometric research, but replacing the role of centralised external exit examinations with public posting of achievement data as a measure of accountability.

The issue with borrowing an argument developed within econometrics research, adapting it to fit a New Public Management perspective, and using it as a policy recommendation, is that the EPA stopped being backed by evidence. In fact, PISA data are available that directly contradict this argument, but they were downplayed by the Directorate for Education and Skills.

Finally, there was the issue of the PISA implicit curriculum. When the OECD suggests that country curricula should be more aligned with 21st century skills it is implicitly conceding that it is not true that PISA is curriculum-free, only that it follows its own. It is unclear what 21st century skills are. Since the contents of PISA are the result of an international effort, it was argued that the PISA curriculum may just be a watered-down synthesis of national curricula.

At present, there is no evidence that reforming curricula so that they are "21st century-ready" improves country outcomes in PISA—with the possible exception of Project Maths in Ireland. It is also unclear whether higher-performing students as measured by PISA have greater achievements in the labour market outcomes as "promised" by the OECD. The fact that different studies found PISA scores to be predictive of a lower risk of school dropout may simply indicate that PISA captures general academic ability rather than skills for life.

RQ2: To what extent do countries follow the OECD education policy advice in their national policies?

The short answer is: to a good extent, but not necessarily because of the OECD. This research question was addressed by complementing cross-country evidence with data from two case studies.

The cross-country analysis returned limited indirect evidence that countries are aligning with OECD recommendations. Inequality is decreasing, but the effect concerns only OECD countries and can only be detected in the correlation between the Strength of the socio-economic gradient and Reading scores. Furthermore, the rate of decrease is very small (3 percentage point in 12 years) and in general equity trends are very stable. The performance of partner countries is increasing, but that of OECD countries is not.

Many experts stated that their countries were more in line than not with the EPA, but a follow-up questionnaire (the PAIQ) revealed that, in a small sample of 15 countries, only five had implemented policies covering more than half of the EPA and only three had privileged equity policies. Given that answering the PAIQ required more time and commitment, it is likely that these findings reflected country contexts more accurately than those from the first expert questionnaire.

Equity was a key aspect of policies in both case studies. The conception of equity in Ireland has traditionally been highly aligned with the OECD's. Many policies addressing socio-economic disadvantage emerged over the years and most of them flowed into DEIS, a cohesive framework covering many EPA categories. Additionally, there have been legislative developments to cater for SEN and immigrant students, but the system is still characterised by religious and gender segregation.

The CFB has a higher level of educational inequality than Ireland. This is due to a fragmentation of the education system that reflects fragmentations in the Community. Because of this, acknowledgement of an OECD goal such as equity happened recently and not homogeneously. The public and private sectors have different visions of what an equitable education system should look like, and attempts to address equity issues in the CFB often did not achieve the intended outcomes.

Both Ireland and the CFB provided examples of policies purporting to be aligned with OECD recommendations which were in fact simply informed by values similar to those of the OECD. One example were the reforms of the Inspectorate, which were driven by greater calls for school accountability in both education systems but fell short of including measures such as public posting of achievement data as advocated by the OECD.

In spite of these findings, which made it explicit for the first time in this research strand that using the OECD or PISA "brands" in policymaking is no guarantee that reforms will actually follow what the OECD says, Irish and Belgian policies were broadly in line with the EPA. This was mostly due to the fact that national policy frameworks such as the Education Act (1998) and the *Décret « Missions »* of 1997 were aligned with the EPA even before PISA, but some post-PISA EPA-aligned reforms were also introduced. Interestingly, the clearest examples of policy alignment (the National Literacy and Numeracy Strategy in Ireland and

the “Pact for an outstanding education” in the CFB) were both recent and had links with prior interventions. This suggests that the influence of the OECD on national policymaking may be increasing, but in an incremental fashion and where it does not create major breaks with existing practice.

In Ireland as well as in the CFB, OECD policy recommendations were only rarely disseminated by OECD staff or external consultants. The role of policy brokers was more often played by the national project managers (the ERC and the ASPE) with the support of public councils such as the National Council for Curriculum and Assessment (NCCA) in Ireland and the *Conseil de l’Education et de la Formation* (CEF) in Belgium. Both the ERC and the ASPE served as policy advisors to the government before PISA and were involved in the administration of other LSAs such as PIRLS and TIMSS, which made them a natural choice as national project managers.

However, a question arises as to whether their expertise may be a double-edged sword. On the one hand, it is certainly preferable for the Secretariat at the OECD and for governments to employ experienced education centres to manage PISA and disseminate policy advice. On the other hand, the independence enjoyed by the ERC and the ASPE—which is necessary for producing high-quality research—may come to the detriment of the EPA when there is disagreement between the OECD and the two research centres on educational matters. The vision and recommendations of the ERC or the ASPE and the OECD were generally in line, but there were instances in which the “voice” of the national project managers was louder and diverged from that of the Secretariat—even though it was still informed by PISA.

The issue is that while the Secretariat and the NPMs have complementary roles on purely administrative matters, they are in competition when it comes to delivering policy advice. The prestige and financial health of the Secretariat depend not only on the quality of its research, but also on the policy relevance of its recommendations, and the same applies to the ERC, the ASPE and probably many other NPMs.

RQ3: What is the evidence that education policies informed by OECD recommendations influenced country performance?

There is little evidence that education policies informed by OECD recommendations influenced country performance. Lack of evidence does not necessarily mean that the advice is not sound, but in the few instances in which it was possible to gather stronger evidence, the outcomes were mixed.

In Chapter 4, it was shown that expanding access to pre-primary education was positively related to PISA outcomes, and it was argued that this may have been because the OECD advice

is itself aligned with solid research evidence. Future PISA cycles will be informative in understanding whether the observed correlation will increase in strength or disappear when new data become available. For now, these findings suggest that PISA might be able to detect policy effects of a similar magnitude, where 1 percentage point increase in an indicator is associated with 3–4 PISA points.

In Chapter 5, it was argued that very few policies in Ireland presented three characteristics simultaneously: namely, being aligned with the EPA; having been implemented effectively; and being sufficiently old that effects on student learning could be evaluated. DEIS was the best example, and evaluations suggested that it had an impact on student outcomes. Preliminary positive results could also be reported for Project Maths.

Because of the few national policy evaluations available, an attempt to evaluate policy impact was made using Willms' (2006) policy evaluation framework. Despite the "universal" nature of many interventions, significant changes were limited almost exclusively to the Science domain and driven by the score increase of the low achievers. If this result captured the effect of the science curriculum, it is unclear why only low achievers benefitted from it. The Strength and the Slope of the socio-economic gradient also remained stable.

In Chapter 6, it was found that Reading performance increased on average because of the higher scores of low achievers, whereas a drop in the performance of high achievers could be observed in Mathematics and Science. Belgian researchers claimed that the increase in Reading scores was due to structural changes in the system and additional instruction, but this does not explain why scores in other domains were not affected by them. Moreover, it was found that EPA-aligned policies addressing systemic inequalities produced little improvement whereas grade repetition—a mechanism which is deeply rooted in the system—worsened.

Overall, no clear-cut policy effects could be identified when applying Willms' framework. Given that this was the case even when national reports were available showing the impact of some policies, it was concluded in the "Limitations and future directions" section that the framework was scarcely helpful when trying to capture policy effects.

Drawing some general conclusions, the clearest instances in which following the EPA was associated with positive outcomes were when the EPA reflected established knowledge in education (as in the case of pre-primary education and DEIS) and was implemented effectively. Lack of wide-scale implementation may be why equity policies in the CFB did not have the desired results—which reminds of the importance of securing the stakeholders' support to achieve policy outcomes.

This is easier said than done. Both in Ireland and in the CFB teachers were consulted before key education reforms, and yet many policies still failed to be integrated into classroom practice. The dynamics of teacher resistance were outside the scope of this research, but it is

clear that involving teachers may only be a pre-condition for successful policy implementation. It should be noted that in neither education system was the teacher professional "continuum" (teacher training, induction and continuing development) substantively improved alongside an increae in teacher accountability. It is possible that teachers felt that the government became more demanding while not providing the conditions for easing policy changes into classroom practice, and that requests for feedback through surveys and public hearings (the National Education Convention of 1993 in Ireland or the *Assises de l'enseignement* in 1995 in the CFB) were too little a step in the right direction.

RQ4: Considering the evidence, how valuable is the OECD education policy advice?

Findings from expert interviews in Ireland and Belgium matched those from previous research whereby, with all its flaws, PISA is a useful instrument to have available. PISA was seen as a 'thermometer', a monitoring or diagnostic tool, as well as a source of valuable data. Experts were aware that PISA can be used 'to support A as well as the opposite of A' (Figazzolo, 2009, p. 28), but for many this was an indication of its quality rather than a limitation: PISA produced 'solid empirical data' (B3), whereas national tests 'can always be suspected of being subjective' (B6). Other experts thought that presenting PISA findings in the most favourable way was part of the political game and not an inherent flaw of PISA, which was considered objective. Even PISA detractors thought that it was convenient to have PISA data to complement national assessment data.

All these opinions concerned what was called in Chapter 1 the "descriptive" or technical aspect of PISA: PISA, PIAAC, TALIS and other surveys are assumed to be a more or less accurate reflection of the state of education across countries and can be used to argue for one's position. This, however, says little about the usefulness of the "normative" component of the OECD work in education policymaking—that is, whether it makes sense to follow the OECD policy advice. A separation between the descriptive and normative sides of PISA (and other OECD surveys) has been generally neglected by education literature, although some authors have pointed out that PISA results and the EPA were two different matters: 'there is no logical connection between, for example, the activities of publishing PISA results […] and providing policy recommendations […] and the assumption that these activities will help actors, particularly at the local level, learn about their own practices and take action to improve education' (Hanberger, 2014, p. 177).

It was shown in this research that the EPA exists and it is linked to PISA, but not necessarily informed by it. It pre-exists and survives PISA, and it encompasses OECD

publications that are only tangentially related to PISA. In line with Hanberger (2014), it was posited that the usefulness of PISA may not coincide with the usefulness of the EPA. Whether the OECD policy advice on educational matters is valuable advice depends on its quality (RQ1), its relevance for national policymaking (RQ2) and its ability to inform effective policy (RQ3). The expected answers to these questions were summarised in a working hypothesis, which is reproduced below from Chapter 2:

> Education policy advice might be vague or inconsistent over time. Despite the powerful narrative about the influence of the OECD in national policymaking, reforms that seem aligned with OECD recommendations may not necessarily informed by them. Of the few policies that make explicit use of the EPA, even fewer might be successfully implemented. Finally, a very small number of implemented policies may have plausibly contributed to (positive or negative) changes in educational outcomes, even when these outcomes are measured by the very assessment that originally informed policymaking.

Unlike what was hypothesised, the EPA is neither vague nor inconsistent over time. Of course it is general, as it has to apply cross-nationally, but what countries should be doing is often explicitly stated. The EPA was also found to be highly consistent over time, so much that even when it *could* draw on the most up-to-date evidence (PISA evidence) it was not necessarily informed by it. Its greater shortcoming, however, lies in its internal incoherence.

There is not one individual policy advice framework, and that produced for this research is a synthesis of multiple publications scattered around the OECD website and outside it. There is no clear dissemination strategy, probably because there is no consensus within the OECD about what mechanisms within education systems are worth focusing on and how they are supposed to interact. This is reflected, for instance, in the ever-changing nature of the education indicators used in PISA. Individual policy recommendations are not strung together in a coherent strategy, even though the OECD often stresses the importance of policy coherence.

The soundness and strength of the recommendations also vary: some are backed by higher-quality evidence than others; some are given more prominence than others; some are theorised more clearly than others; and only a few show these three characteristics together (e.g., the negative impact of streaming policies). The OECD openly states that PISA is a work in progress, but to some extent each cycle feels like a new pilot project. Nevertheless, the 2015 assessment framework is a clear attempt to introduce some consistency in the educational vision of the OECD.

It is true that not all national reforms acknowledging the OECD goals and values follow the EPA, and few instances of policy enactment (when a policy is both aligned with the EPA

and acknowledges the contribution of the OECD or PISA) were found. While it is unclear on a global scale how many policies are in line with the EPA, many instances of policy alignment were observed in Ireland and the CFB. Both national education frameworks were already in substantive agreement with the EPA before PISA, but even greater policy convergence was reached from the mid-2000s. This suggests that PISA may have been an effective method for delivering the EPA in these receptive jurisdictions.

Many interventions "broke down" at the implementation level and were not integrated into classroom practice. There is also very little evidence that following the EPA is beneficial for the quality and the equity of the education systems under investigation. EPA-informed policies may or may not work. Moreover, the limitations of PISA in capturing changes below 20 or even 30 points reliably, as well as the inability of Willms' (2006) evaluation framework to detect policy effects, suggest that it is still too soon to base policy advice on PISA trends.

To conclude, Table 7.1 reviews the initial argument statement provided in Chapter 2 in light of the knowledge gathered during this research.

Table 7.1: A revision of the initial thesis statement in light of the knowledge gathered during this research

Initial argument	Final argument
Is it possible that despite the use of data from PISA within political and scientific circles, OECD recommendations are currently of little educational value, because they are not solidly grounded empirically; they provide vague policy direction; advice is not followed; and even when it is followed it is not associated with measurable changes?	The educational value of OECD recommendations varies. Some are solidly grounded, but other are not. Policy direction is provided, but it does not reflect any clear strategy. There is some evidence that OECD advice is increasingly being followed after the first two or three PISA cycles, but very limited evidence that following it is associated with measurable changes.

As mentioned earlier, one natural avenue for future research would be to update the OECD policy advice and continue to track cross-sectional data over the next PISA cycles, to see whether more countries will start improving and the increase experienced by a range of countries will be sustained. If the OECD policy databases become more thorough and granular, these data could eventually be crossed to find more convincing (albeit still indirect) evidence of EPA effects on country outcomes. Additional case studies could be investigated in parallel, as time and resource constraint limited the current analysis to two.

A different but equally compelling future direction would be to build upon one of the key findings from this research—the positive effects of pre-primary education on student achievement. Like the OECD's thematic reviews, such research would focus on comparing

the effects of one type of policy across many countries. Unlike the OECD reviews, that hypothetical study would employ two methodological improvements that were used in this thesis. One would be to pay greater attention to policy alignment and convergence, including the disconnect between acknowledgement and alignment, on the one hand, and policy intention and implementation (or integration), on the other hand. The second improvement would be to complement correlational analyses with multilevel regressions that modelled more explicitly and accurately the effects of time, so as to provide more robust evidence that *introducing* or *altering* early childhood policies (and not simply having versus not having them) can have an effect on educational outcomes.

Volume 2: Appendices and references

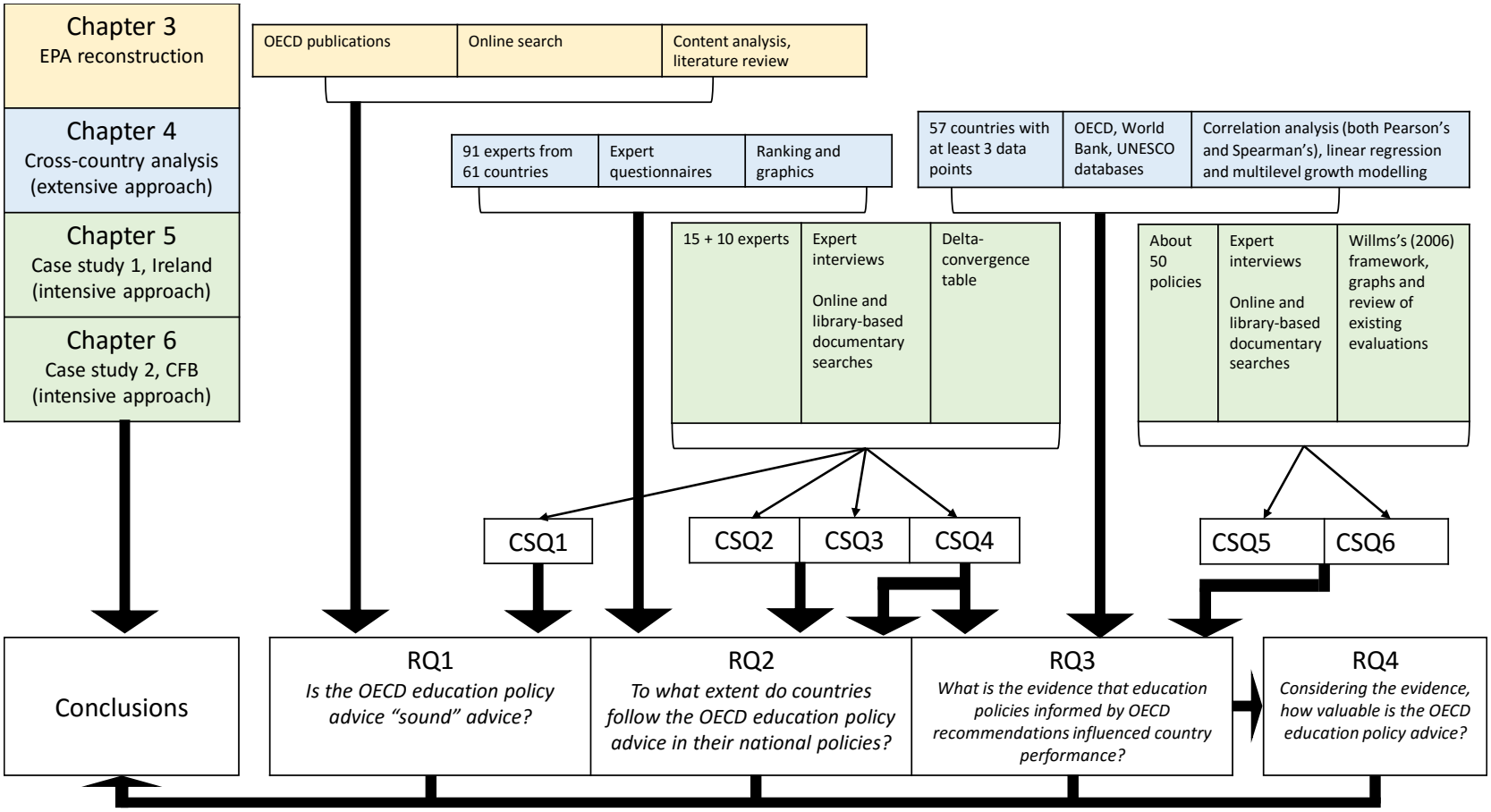
Volume 2 of 2

Table of contents (Volume 2)

List of tables (Volume 2)

List of figures (Volume 2)

Appendix 1: Research design



Appendix 2: A comparison of the education systems in Ireland and in the CFB

1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
Age	**ISCED**	**Education (1)**	**Enseign. (1)**	**Edu. (2)**	**Ens. (2)**	**Curriculum**	**Continuum pédagogique**	**Class/Year**	**Année**	**Qual.**	**Certification**
17/18	3	Post-primary	Secondaire	Senior Cycle	3e degré	LC	Tr (G,T,A) / Q (T,A,P)	6th year	6e	LCE / LCA / LCVP	CESS
16/17	3	Post-primary	Secondaire	Senior Cycle	3e degré	LC	Tr (G,T,A) / Q (T,A,P)	5th / 6th year	5e	LCE / LCA / LCVP	
15/16	3	Post-primary	Secondaire	Senior Cycle	2e degré	LC / TY	Tr (G,T,A) / Q (T,A,P)	TY / 5th year	4e		CES2D
14/15	2	Post-primary	Secondaire	Junior Cycle	2e degré	JC	Tr (G,T,A) / Q (T,A,P)	3rd year	3e	JC	
13/14	2	Post-primary	Secondaire	Junior Cycle	1er degré	JC	3e étape, C/D/S	2nd year	2e		CEB / CE1D
12/13	2	Post-primary	Secondaire	Junior Cycle	1er degré	JC	3e étape, C/D/S	1st year	1re		CEB
11/12	1	Primary	Fondamental		Primaire	Primary	2e étape, 2e cycle	6th class	6e		CEB
10/11	1	Primary	Fondamental		Primaire	Primary	2e étape, 2e cycle	5th class	5e		
9/10	1	Primary	Fondamental		Primaire	Primary	2e étape, 1er cycle	4th class	4e		
8/9	1	Primary	Fondamental		Primaire	Primary	2e étape, 1er cycle	3rd class	3e		
7/8	1	Primary	Fondamental		Primaire	Primary	1re étape, 2e cycle	2nd class	2e		
6/7	1	Primary	Fondamental		Primaire	Primary	1re étape, 2e cycle	1st class	1re		
5/6	1	Primary	Fondamental		Maternel	Primary	1re étape, 2e cycle	Senior infant	3e		
4/5	1	Primary	Fondamental		Maternel	Primary	1re étape, 1er cycle	Junior infant	2e		
3/4	0	Early childhood	Fondamental		Maternel		1re étape, 1er cycle	Early Start	1re		

Sources: Ministère de la Communauté française (2008); Department of Education and Science (2004); online resources

The table in the previous page compares the education systems in Ireland and in the CFB. With the exception of columns 1 and 2 (*Age* and *ISCED*), which are common to both jurisdictions, the rest of the table alternates Irish categories (columns 3, 5 … 11) to Belgian categories (columns 4, 6 … 12). Notes and explanations are given below.

Column	Notes
1	Education is compulsory in Ireland from the age of 6 to 16, and in the CFB from the age of 6 to 18 (though the last two years can be part-time)
2	The "International Standard Classification of Education", or ISCED (UNESCO-UIS, 2012) is a 'standard framework used to categorise and report cross-nationally comparable education statistics' (p. iii). ISCED level 0 corresponds to early childhood education. ISCED 1 is primary education, 2 is lower secondary and 3 is upper secondary education.
3–6	One way to organise education systems is in terms of education levels. For instance, one can refer to 'first-level education' in Ireland or '*enseignement fondamental*' in the CFB (columns 3 and 4). Alternatively, one can speak of 'primary education' or '*enseignement primaire*' to refer broadly to the same period (columns 5 and 6). Note, however, that the Belgian '*enseignement fondamental*' (column 4) includes the final year of ISCED 0, and that the '*enseignement maternel*' extends to the age of 5/6 (column 6), roughly corresponding to kindergarten in some countries. Another difference is the organisation of secondary education. In Ireland there are a Junior and a Senior Cycle that broadly correspond to ISCED 2 and 3, the CFB has three two-year-long stages (*degrés*, column 6). The most relevant for this research is the first stage of lower-secondary education (*premier degré du secondaire*), corresponding to the ages 12 to 14 for a child who did not repeat any year, which is common practice in the CFB (recent data show that only about 65% of the pupils enter this stage at the correct age; see Baye et al., 2014). The *1er degré* was the object of many reforms in the last few years, both in terms of syllabus and instruction and in terms of enrolment procedures.
7–8	A second way to organise education systems is in terms of taught curriculum, or *continuum pédagogique* (columns 7 and 8). In Ireland, there is a common primary curriculum followed by a common Junior Certificate (JC) curriculum, which leads to the qualification of the same name (shown in column 13). In upper-secondary education, Irish students follow the Leaving Certificate (LC) curriculum (leading to the Leaving Certificate Established, LCE), or vocational syllabi such as the Leaving Certificate Applied (LCA) programme or the Leaving Certificate Vocational Programme (LCVP). More than half of Irish students do not enter LC directly from JC, but prefer to enrol in an optional Transition Year (TY) where they can pursue non-academic interests and 'develop a range of transferable critical thinking and creative problem-solving skills' through 'active and experiential' learning (Department of Education and Science, 2006; see also the work of Clerkin, 2013, for a recent evaluation). There is no common curriculum in the CFB, although the introduction of the Core and Final Skills in 2001 was an attempt to standardise the educational experiences of Belgian pupils. To give a sense of structure and continuity while abiding by the constitutional freedom of education, the pedagogical continuum in the CFB is organised in terms of legs (*étapes*) and cycles (column 8). Over the years, many policies have tried to reduce grade repetition by prohibiting the practice within cycles or and sometimes *étapes*, but schools are allowed to add one supplementary year (*année complémentaire*) per *étape* to the educational career of pupils to account for individual learning rhythms. In the *3e étape*, which corresponds to the *1er degré*, pupils can follow a common curriculum (C), a differentiated curriculum (*année différenciée*, D) leading to the (CEB)—if they failed to attain it at the end of primary education (see column 14)—or one of the supplementary years (S).

	Starting from the *2e degré*, students are streamed into two main branches: *Transition* (Tr), the academic path leading students to higher education; and *Qualification* (Q), the vocational option. Each branch is subdivided into three tracks determining the school curriculum. The *enseignement de transition* is divided into *Général* (G), *Technique* (T) and *Artistique* (A). The two latter tracks are available also in the *enseignement de qualification*: the idea is to provide two different "flavours"—one more, one less academically-oriented—to technical and artistic education. Moreover, vocational education also has the *Professionnel* (P) route leading to lower-skill occupations.
9, 10	A third way to organise education systems is in terms of grades. It is customary both in Ireland and in the CFB to restart the count at each level of education (columns 9–10). At the primary level, one generally speaks of 1st–6th Class in Ireland and *1e–6e année primaire* (or 1P–6P) in the CFB. At the secondary level, Irish students go from 1st to 3rd year and then they move either to Transition Year or directly to 5th Year (a 4th year does not exist); this means that the Irish 5th and 6th Years gather students from two age groups: those that did and those that did not do a TY. Likewise, it means that the Senior Cycle can last two or three years depending on this choice. Because of TY, the PISA cohort in Ireland is enrolled either in TY or in the 5th Year. In Belgian documents, the number of the year is generally followed by a letter indicating the curriculum or track. For instance, 1C or *première commune* refers to the first year of lower-secondary education following the common curriculum, 1D or *première différenciée* is the first year of lower-secondary education in the differentiated curriculum, and so on. Because of grade retention, the PISA cohort in the CFB may be enrolled anywhere between the *1e* and the *4e années* of secondary education.
11, 12	Neither the JC nor the various LC options are compulsory in Ireland, though they are taken by most students. In the CFB, the CEB is compulsory at the end of primary education (6P). Pupils who attain the CEB enter secondary education in 1C, otherwise they enter 1D or 1S and resit the examination one year later. If they are unsuccessful, they continue in 2D/S. At the end of *1er degré*, pupils sit the CE1D, which is compulsory since 2013–2014. Remedial classes similar to the D and S routes are organised in case of failure. The CES2D is not compulsory, but the CESS is.

Appendix 3: Additional tables and figures

Chapter 1

Table 1.A: Members of the expert groups

Name	Domain	Country	2000	2003	2006	2009	2012	TOT
Dominique Lafontaine	R	Belgium (Fr)	1	1	1	1	1	5
Stan Jones	R	Canada	1	1	1	0	0	3
Pirjo Linnakylä	R	Finland	1	1	1	1	1	5
Jean-Francois Rouet	R	France	0	0	0	1	1	2
Martine Rémond	R	France	1	1	1	0	0	3
Wolfgang Schneider	R	Germany	1	0	0	0	0	1
Wolfgang Schnotz	R	Germany	0	0	0	1	1	2
Ryo Watanabe	R	Japan	1	0	0	0	0	1
Sachiko Adachi	R	Japan	0	0	0	1	1	2
Minwoo Nam	R	Korea	0	0	0	1	1	2
John de Jong	R	Netherlands	1	1	1	1	1	5
Eduardo Vidal-Abarca	R	Spain	0	0	0	1	1	2
Alan Davies	R	UK (Eng)	1	1	1	1	1	5
Charles Alderson	R	UK (Eng)	0	0	0	1	1	2
Irwin Kirsch	R	USA	1	1	1	1	1	5
John Guthrie	R	USA	0	0	0	1	1	2
Marilyn Binkley	R	USA	1	1	1	0	0	3
Caroline Bardini	M	Australia	0	0	0	0	1	1
Kaye Stacey	M	Australia	0	0	0	0	1	1
Peter Schüller	M	Austria	1	0	0	0	0	1
Mogens Niss	M	Denmark	1	1	1	1	1	5
Werner Blum	M	Germany	0	1	1	1	1	4
Sean Close	M	Ireland	1	1	0	0	0	2
Raimondo Bolletta	M	Italy	1	0	0	0	0	1
Toshikazu Ikeda	M	Japan	0	0	0	0	1	1
Yoshinori Shimizu	M	Japan	0	1	1	1	0	3
Kyung Mee Park	M	Korea	1	1	0	0	0	2
Jan de Lange	M	Netherlands	1	1	1	1	0	4
Zbigniew Marciniak	M	Poland	0	1	1	1	1	4
Valdimir Burjan	M	Slovak Republic	0	1	0	0	0	1
Luis Rico	M	Spain	0	1	0	0	0	1
Maria Luisa Moreno	M	Spain	1	0	0	0	0	1
Martin Ripley	M	UK (Eng)	0	0	0	0	1	1
Joan Ferrini-Mundy	M	USA	0	0	0	0	1	1
John Dossey	M	USA	0	1	1	1	0	3
Mary Lindquist	M	USA	0	1	0	0	0	1
Solomon Garfunkel	M	USA	0	0	0	0	1	1

Thomas Romberg	M	USA	1	0	0	0	0	1
William Schmidt	M	USA	0	0	0	0	1	1
Peter Fensham	S	Australia	1	1	1	1	1	5
Robert Laurie	S	Canada	0	0	1	0	0	1
Andrée Tiberghien	S	France	0	0	0	1	1	2
Pierre Malléus	S	France	0	0	1	0	0	1
Manfred Prenzel	S	Germany	1	1	0	1	1	4
Michelina Mayer	S	Italy	0	0	1	0	0	1
Yasushi Ogura	S	Japan	0	0	1	1	1	3
Donghee Shin	S	Korea	1	1	0	0	0	2
Svein Lie	S	Norway	1	1	1	1	1	5
Ewa Bartnik	S	Poland	0	0	1	0	0	1
Paulina Korsnakova	S	Slovak Republic	0	0	1	0	0	1
Raul Gagliardi	S	Switzerland	1	1	0	0	0	2
Robin Millar	S	UK (Eng)	0	0	1	0	0	1
Wynne Harlen	S	UK (Scot)	1	1	0	0	0	2
Elizabeth Stage	S	USA	1	1	0	0	0	2
Rodger Bybee	S	USA	0	0	1	1	1	3
Senta Raizen	S	USA	1	1	0	0	0	2
Pascal Bressoux	Q	France	-	-	-	1	0	1
Eckhard Klieme	Q	Germany	-	-	-	1	1	2
Ludger Wößmann	Q	Germany	-	-	-	1	0	1
Yin Cheong Cheng	Q	Hong Kong	-	-	-	1	0	1
Eduardo Backhoff	Q	Mexico	-	-	-	0	1	1
Fons van de Vijver	Q	Netherlands	-	-	-	0	1	1
Jaap Scheerens	Q	Netherlands	-	-	-	1	1	2
Ying-yi Hong	Q	Singapore	-	-	-	0	1	1
David Kaplan	Q	USA	-	-	-	1	1	2
Henry Levin	Q	USA	-	-	-	1	1	2

R = Reading, M = Mathematics, S = Science, Q = Background Questionnaire. *Sources*: PISA Assessment Frameworks 2000–2012

Table 1.B: What country agencies are represented in the PISA Governing Board and as National Project Managers?

ID	NAME	OECD	PISA Governing Board	National Project Manager	Same agency?
1	Australia	yes	Ministry of Education	ACER	no
2	Austria	yes	Ministry of Education	BIFIE	no
3	Belgium (Fl)	yes	Ministry of Education	University of Ghent	no
4	Belgium (Fr)	yes	ASPE	ASPE	yes
5	Canada	yes	Ministry of Education	Ministry of Education	yes
6	Chile	yes	Agency of Quality of Education	Agency of Quality of Education	yes
7	Czech Republic	yes	Czech School Inspectorate	Czech School Inspectorate	yes
8	Denmark	yes	Ministry of Education	KORA	no
9	Estonia	yes	Ministry of Education	Foundation Innove	no
10	Finland	yes	Ministry of Education	University of Jyväskylä	no
11	France	yes	Ministry of Education	Ministry of Education	yes
12	Germany	yes	Ministry of Education	ZIB	no
13	Greece	yes	Harokopio University	Harokopio University	yes
14	Hungary	yes	Ministry of Education	Ministry of Education	yes
15	Iceland	yes	Ministry of Education	Namsmatsstofnun	no
16	Ireland	yes	ERC	ERC	yes
17	Israel	yes	Ministry of Education	Ministry of Education	yes
18	Italy	yes	INVALSI	INVALSI	yes
19	Japan	yes	NIER	NIER	yes
20	Korea	yes	KICE	KICE	yes
21	Luxembourg	yes	Ministry of Education	Ministry of Education	yes
22	Mexico	yes	INEE	INEE	yes
23	Netherlands	yes	Ministry of Education	Cito	no
24	New Zealand	yes	Ministry of Education	Ministry of Education	yes
25	Norway	yes	Ministry of Education	University of Oslo	no
26	Poland	yes	Ministry of Education	IBE	no
27	Portugal	yes	IAVE	IAVE	yes
28	Slovak Republic	yes	NUCEM	NUCEM	yes

29	Slovenia	yes	Educational Research Institute	Educational Research Institute	yes
30	Spain	yes	Ministry of Education	Ministry of Education	yes
31	Sweden	yes	Ministry of Education	Mid Sweden University	no
32	Switzerland	yes	Ministry of Education	Other	no
33	Turkey	yes	Ministry of Education	Ministry of Education	yes
34	UK (Eng)	yes	Ministry of Education	RM Results	no
35	UK (Scot)	yes	Ministry of Education	NFER	no
36	United States	yes	Ministry of Education	Ministry of Education	yes
37	Albania	no	Ministry of Education	National Agency of Examinations	no
38	Algeria	no	Ministry of Education	Ministry of Education	yes
39	Argentina	no	Ministry of Education	Ministry of Education	yes
40	Brazil	no	INEP	INEP	yes
41	Bulgaria	no	Ministry of Education	Ministry of Education	yes
42	China	no	Ministry of Education	Ministry of Education	yes
43	Colombia	no	ICFES	ICFES	yes
44	Costa Rica	no	Ministry of Education	Ministry of Education	yes
45	Croatia	no	National PISA Centre	National PISA Centre	yes
46	Dominican Republic	no	Ministry of Education	Ministry of Education	yes
47	Georgia	no	NAEC	NAEC	yes
48	Hong Kong SAR	no	Chinese University of Hong Kong	Chinese University of Hong Kong	yes
49	Indonesia	no	Ministry of Education	Ministry of Education	yes
50	Jordan	no	Ministry of Education	Ministry of Education	yes
51	Kazakhstan	no	MESRK	MESRK	yes
52	Kosovo	no	Ministry of Education	Ministry of Education	yes
53	Latvia	no	University of Latvia	University of Latvia	yes
54	Lebanon	no	Ministry of Education	Ministry of Education	yes
55	Lithuania	no	Ministry of Education	National Examination Center	no
56	Macao SAR	no	Ministry of Education	University of Macau	no
57	Macedonia	no	NAEEPM	NAEEPM	yes
58	Malaysia	no	Ministry of Education	Ministry of Education	yes
59	Malta	no	University of Malta	Ministry of Education	no

60	Moldova	no	Ministry of Education	Ministry of Education	yes
61	Montenegro	no	Examination Centre	Examination Centre	yes
62	Peru	no	Ministry of Education	Ministry of Education	yes
63	Qatar	no	Ministry of Education	Ministry of Education	yes
64	Romania	no	CNEE	CNEE	yes
65	Russian Federation	no	Russian Academy of Education	Russian Academy of Education	yes
66	Serbia	no	Ministry of Education	University of Belgrade	no
67	Singapore	no	Ministry of Education	Ministry of Education	yes
68	Taiwan	no	Ministry of Education	National Chiao Tung University	no
69	Thailand	no	IPST	IPST	yes
70	Trinidad and Tobago	no	Ministry of Education	Ministry of Education	yes
71	Tunisia	no	CNIPRE	CNIPRE	yes
72	United Arab Emirates	no	Ministry of Education	Ministry of Education	yes
73	Uruguay	no	ANEP	ANEP	yes
74	Vietnam	no	Ministry of Education	Ministry of Education	yes

Source: OECD website. When a country had more than one delegate in the PGB or as NPM but one institute featured in both groups, that institute was reported in both columns. Note that even when the Ministry of Education does appear in the list, some institutes may be dependent from it.

Chapter 2

Table 2.A: the draft model of multi-agent practical reasoning applied to one OECD argument

Means (EPA)	Goals	Means-goals	Values	Circumstances
Ensure that the quality of pre-primary education is high	Improve performance	If quality is high, performance will improve, because quality improves performance.	Performance ought to be valued	-
Ensure that low-SES families can access pre-primary education nearby	Improve performance and equity	If l low-SES families can access pre-primary education equity and performance will improve, because access to pre-primary education improves performance	Performance and equity ought to be valued	'Pre-primary enrolment rates are growing faster among advantaged students'
Ensure that low-SES families know about pre-primary programmes	Improve performance and equity	If low-SES families know about pre-primary programmes	Performance and equity ought to be valued	'Pre-primary enrolment rates are growing faster among

		equity and performance will improve, because more information will lead to more access and therefore to higher equity and performance		advantaged students'
Invest in pre-primary education	Improve efficiency	Investment will increase economic returns (efficiency) because performance is linked to growth and increasing access and quality will increase performance	Efficiency and economic returns ought to be valued	-

Source: Own interpretation of an OECD policy recommendation (2014c, p. 4).

Table 2.B: the revised practical reasoning model applied to an OECD argument from the Education Policy Outlook for Australia.

Consider this passage from the *Education Policy Outlook* for Australia:

> '*Fair and inclusive policies* contribute to an equitable education system. […] Australia has a high degree of school choice which, if not well targeted, can contribute to segregation of students. Around 96% of students attend a school that competes for enrolments with at least one other school. Evidence shows that school choice, along with other factors, may undermine equity in the education system by segregating students into schools based on their socio-economic background. It is important to ensure that there are mechanisms to mitigate this negative effect.' (OECD, 2013b, p. 6, emphasis in the text)

Problem	**Evidence**
Australia has a high degree of school choice which, if not well targeted, can contribute to segregation of students	Probably national statistics
Goals and values	
Equity in the education system	
Circumstances	
Around 96% of students attend a school that competes for enrolments with at least one other school	Probably national statistics
Means-goal	
If Australia limited school choice or counteracted its effects there would be more equity	'Fair and inclusive policies contribute to an equitable education system' 'Evidence shows that school choice, along with other factors, may undermine equity in the education system by segregating students into schools based on their socio-economic background'
EPA	

It is important to ensure that there are mechanisms to mitigate the effect of school choice on segregation	

Source: : Own interpretation of an OECD policy recommendation (2013b, p. 6).

Table 2.C: Sample email and integrated questionnaire sent to national education experts

Email subject: Brief questionnaire on education policies and OECD

Dear expert,

As part of my doctoral thesis at the University of Durham *'The relationship between international large-scale assessments and national education reforms'*, I am asking education policy experts around the world to give an overall anonymous rating of their government's policies over the last decade.

I was wondering if you could help by answering two brief questions (below, all personal information will be deleted after replying to your email):

What is your country?
To what extent have your government's education policies been in line with the OECD's advice, over the last decade? (for a summary of OECD's policy advice, see http://community.dur.ac.uk/cesare.aloisi/OECD_recommendations.html)

a. Not at all
b. Very slightly
c. To a modest extent
d. Quite a lot
e. Completely

Thank you very much for your time, your contribution is extremely appreciated. If you think any other colleagues/contacts are able to provide additional information by completing the same questionnaire, I would be grateful if you could forward this message to them to help collect more data for my PhD.

Table 2.D: The OECD policy advice implementation questionnaire (paper version)

Which areas did your country most effectively target through education policies, interventions and reforms, in the last decade? Please tick each relevant box:

	Educational time		
1	**Tick**	**Area of reform**	**Interventions in this area may include…**
a		Increasing overall time spent in education	expanding access to pre-primary education, targeting enrolment or changing the age of entry into school
b		Increasing instruction time	targeting lateness, attendance, by introducing full-day schooling, or by increasing the time spent in classes
	Equality of opportunity to learn		
2	**Tick**	**Area of reform**	**Interventions in this area may include…**
a		Reducing disabled or immigrant students' segregation	by reviewing the sorting of pupils in institutions, or by better allocating special education teachers where they are most needed
b		Reducing the gender gap	for instance, by promoting subjects/career paths that are traditionally dominated by one gender to boys or girls
c		Reducing the streaming / grouping / stratification of students	by reviewing the academic or vocational paths, reducing the transfer of students to different schools, reviewing retention policies or ability grouping within classes

d		Reducing school competition	by changing school incentives to take in the 'best' students or by tackling the elite status of private institutions
e		Helping socio-economically deprived students	by providing additional instructions, free learning resources, fee waivers, cash transfers/vouchers or a modified curriculum
	Quality of provision		
3	**Tick**	**Area of reform**	**Interventions in this area may include…**
a		Strengthening the teaching profession	qualifications required, initial teacher training, allocation of excellent teachers/principals where they are needed the most, retention of teachers in the most challenging schools, teacher pay, in-service training/professional development, career paths
b		Allocating resources more efficiently or in a more equitable way	by changing the share of spending towards public and private schools, by targeting socio-economic depressed areas or acting upon the evaluation of intervention programmes
c		Pairing more autonomy with accountability measures	by increasing school autonomy in budget, curriculum and/or assessments **while at the same time** holding schools accountable for their results by public posting of results, external monitoring of standards
d		Promoting cooperation between parents, teachers, principals and schools	through collaborative programmes, by increasing parental involvement in school matters and in school governance, by facilitating/incentivising school-based research or the sharing of good practice
	Student performance		
4	**Tick**	**Area of reform**	**Interventions in this area may include…**
a		Helping low-performing students	through feedback and appraisals for students, or by establishing early-warning mechanisms and providing a modified curriculum or additional instructional support for struggling students
b		Standardising the educational outcomes	by introducing central, external exit examinations for qualification purposes
c		Reviewing curricular priorities	by introducing curricular that are better aligned with students' interests and 21st century skills, or by examining curricular choices in the light of the performance of other countries

Chapter 3

Tables

Table 3.A: Policy lever and policy themes in the Education Policy Outlook Reforms Finder

Lever	Theme
Equity and quality	Disadvantaged students
	Investing early on
	System-level policies that promote or hinder equity
Evaluation and assessment to improve student outcomes	Student assessment
	School evaluation
	System evaluation
Funding	Economic resources in education
	Use of resources
Governance	Education priorities
	Organisation of decision-making process
Preparing students for the future	Quality of secondary
	Vocational education and training
	Quality of tertiary
	Transition between school and work
School improvement	Learning environments
	School leadership
	Teachers

Source: Education Policy Outlook Reforms Finder

Table 3.B: Educational areas available on the Education GPS website

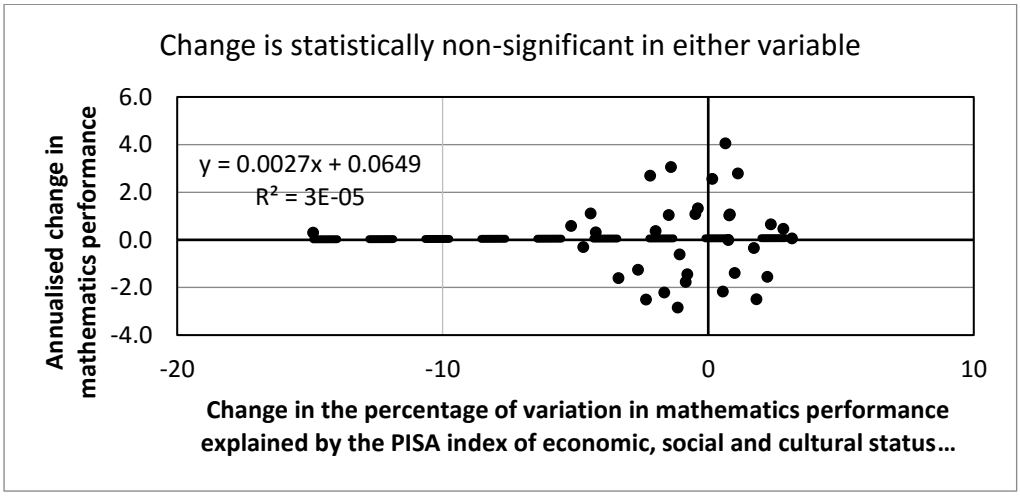
Area	Aspect
Access & Participation	Access & Participation
	Career guidance
Attainment	Attainment
Economic & Social outcomes	Social & Health outcomes
	Innovation
	Labour market outcomes
	Regional development
	Public returns
Equity	Socio-economic status
	Migrant background
Evaluation & Quality assurance	Teacher appraisal
	System evaluation
	School leader appraisal
	Student assessment
	School evaluation
Finance & Funding	Sources of funding
	Uses and allocation of public funds
Internationalisation	Student mobility
	Cross-border quality assurance
Learning environment	Learning time & Disciplinary climate
	Teacher practices
	Class size & Student-teacher ratio
Organisation & Governance	Parental involvement
	School choice
	Public & Private stakeholders
	School autonomy
Research & Innovation in education	Research & Innovation in education

School leadership	School leader working conditions
	School leader preparation and development
	School leader employment
Skills	Skills
Teachers	Teacher initial education
	Teacher working conditions
	Teacher professional development
	Teacher employment
Trends shaping education	Trends shaping education

Source: Education GPS

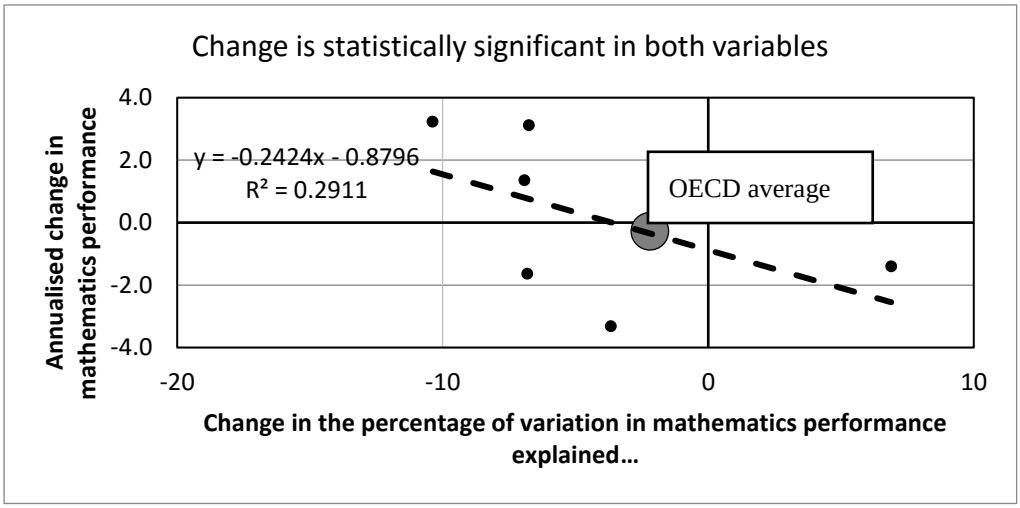
Figures

Figure 3.A: Non-significant change between 2003 and 2012 in the strength of the socio-economic gradient and annualised mathematics performance



Source: own elaboration using data from the PISA 2012 Database, Figure II.2.12, available from http://dx.doi.org/10.1787/888932964813

Figure 3.B: Statistically significant change between 2003 and 2012 in the strength of the socio-economic gradient and annualised mathematics performance



Source: own elaboration using data from the PISA 2012 Database, Figure II.2.12, available from http://dx.doi.org/10.1787/888932964813

Chapter 4

Table 4.A: The country sample used in this study

Country	00	03	06	09	12	Country	00	03	06	09	12
Albania		N/A	N/A			*Korea*					
Argentina		N/A				*Latvia*					
Australia						*Liechtenstein*					
Austria				N/A		*Lithuania*	N/A	N/A			
Azerbaijan	N/A	N/A			N/A	*Luxembourg*	N/A				
Belgium						*Macao*	N/A				
Brazil						*Malaysia*	N/A	N/A	N/A		
Bulgaria		N/A				*Mexico*					
Canada						*Montenegro*	N/A	N/A			
Chile		N/A				*Netherlands*	N/A				
Taipei	N/A	N/A				*New Zealand*					
Colombia	N/A	N/A				*Norway*					
Costa Rica	N/A	N/A	N/A			*Peru*		N/A	N/A		
Croatia	N/A	N/A				*Poland*					
Czech Rep.						*Portugal*					
Denmark						*Qatar*	N/A	N/A			
Dubai	N/A	N/A	N/A			*Romania*		N/A			
Estonia	N/A	N/A				*Russian Federation*					
Finland						*Serbia*	N/A	N/A			
France						*Shanghai*	N/A	N/A	N/A		
Germany						*Singapore*	N/A	N/A	N/A		
Greece						*Slovak Republic*	N/A				
Hong Kong						*Slovenia*	N/A	N/A			
Hungary						*Spain*					
Iceland						*Sweden*					
Indonesia						*Switzerland*					
Ireland						*Thailand*					
Israel		N/A				*Tunisia*	N/A				
Italy						*Turkey*	N/A				
Japan						*UAE (w/o Dubai)*	N/A	N/A	N/A		
Jordan	N/A	N/A				*United Kingdom*	N/A	N/A			
Kazakhstan	N/A	N/A	N/A			*United States*			(R only)		
Kyrgyzstan	N/A	N/A			N/A	*Uruguay*	N/A				

The columns of the table represent data availability for each PISA administration, which takes into consideration the exclusion of some specific countries and years from trend comparisons because of anomalies in their test administration and sampling (OECD, 2010d).

Table 4.B: The reliabilities of PISA scales between 2000 and 2012

	2000	2003	2006	2009	2012
Reading	0.93	0.848	0.891	0.921	0.888
Mathematics	0.90	0.918	0.892	0.882	0.914
Science	0.90	0.843	0.920	0.896	0.885

Source: PISA technical reports 2000–2012

Table 4.C: Model-building approach to analyse trends over time in strength and slope of the socio-economic gradient

	Model					
Component	A	B	C	D	E	F
Intercept (random)	✓	✓	✓	✓	✓	✓
Time (fixed)		✓	✓	✓	✓	✓
Time (random)			✓	✓	✓	✓
Time (quadratic)				✓		
OECD (dummy)					✓	✓
*Time*OECD*					✓	✓
"X" variable (GDP, pre-primary, equity etc.)						✓

Table 4.D: A comparison of different models to map PISA trends over time

	Reading					
	Model A		**Model B**		**Model C**	
	Estimate	*Error*	*Estimate*	*Error*	*Estimate*	*Error*
(intercept)	463.34***	(6.67)	457.35***	(6.85)	454.61***	(7.95)
Time	-		2.48***	(0.56)	3.34***	(0.77)
σ^2_{u0}	2891	(2037–4102)	2931	(2067–4157)	4012	(2785–5780)
σ^2_{u1}	-		-		20.45	(9.51–43.97)
σ_{u01}	-		-		-233.66	
σ^2_e	153.8	(126.21–187.42)	140.17	(114.98–170.90)	104.31	(81.95–132.78)
Log L	-1176.264		-1167.038		-1149.466	
(df), p diff			(1), <0.0001		(2), <0.0001	
	Mathematics					
	Model A		**Model B**		**Model C**	
	Estimate	*Error*	*Estimate*	*Error*	*Estimate*	*Error*
(intercept)	468.02***	(7.40)	465.83***	(7.50)	462.94***	(8.48)
Time	-		1.25**	(0.63)	2.44***	(0.93)
σ^2_{u0}	3587	(2536–5075)	3603	(2547–5097)	4665	(3278–6638)
σ^2_{u1}	-		-		40.09	(23.14–69.45)
σ_{u01}	-		-		-315.84	
σ^2_e	94.16	(75.57–117.32)	92.31	(74.03–115.09)	44.45	(32.73–60.36)
Log L	-989.9554		-988.0101		-959.0912	
(df), p diff			(1), 0.0486		(2), <0.0001	
	Science					
	Model A		**Model B**		**Model C**	
	Estimate	*Error*	*Estimate*	*Error*	*Estimate*	*Error*
(intercept)	472.31***	(6.78)	469.73***	(6.81)	469.33***	(7.04)
Time	-		2.45***	(0.67)	2.71*	(0.83)
σ^2_{u0}	3009	(2128–4254)	3009	(2129–4254)	3239	(2288–4586)
σ^2_{u1}	-		-		29.23	(16.87–50.63)
σ_{u01}	-		-		-119.77	
σ^2_e	57.78	(44.86–74.43)	52.46	(40.68–67.64)	24.66	(17.02–35.72)
Log L	-805.1779		-798.8580		-784.0280	

(df), p diff			(1), <0.001		(2), <0.0001	

****p* <0.01, ***p* <0.05, **p* <0.1

Model A is a null model assuming no time effects but with random intercept (γ_{00}). Model B is the random intercept, fixed slope (γ_{10}) model; and Model C is the random intercept, random slope model. All models have been estimated using restricted maximum likelihood using the "nlme" package (version 3.1-117, Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2015) in the statistical environment R (R Core Team, 2015). Note that the analysis of variance (Log L) between A and B was carried out after estimating the two models using standard maximum likelihood, because B has an additional fixed effect. Standard errors are for the fixed effects and confidence intervals (at the 95% confidence level) for random effects. Confidence intervals are preferred as an indication of error on the variance component because these latter are non-symmetrical. It is acknowledged that variance components often do not have a normal distribution and intervals calculated in this fashion can only be an approximation (Bates, 2009; Goldstein, 2011).

Table 4.E: Between-country variance as a function of time

	2000	**2003**	**2006**	**2009**	**2012**
Reading	4012	3565.13	3159.16	2794.09	2469.92
Mathematics		4665	4073.41	3562	3130.77
Science			3239	3028.69	2876.84

Modelled using information from Model C in Table 4.D above.

Table 4.F: Initial level (β_0) and average change (β_1) of the Strength and Slope of the socio-economic gradient in different countries

	Reading				**Mathematics**				**Science**			
	Strength		**Slope**		**Strength**		**Slope**		**Strength**		**Slope**	
ID	β_0	β_1	β_0	β_1	β_0	β_1	β_0	β_1	β_0	β_1	β_0	β_1
ARG							47.53	-6.62			41.57	-5.16
DEU	21.56	-1.69	53.98	-3.59	22.59	-1.98						
FIN			24.13	2.26								
FRA			43.66	3.97			46.46	4.65				
GBR	17.56	-1.49			18.46	-2.19						
ISL			21.04	2.3								
ISR					8.95	2.8						
JOR	18.22	-3.37	51.18	-7.84			39.23	-5.43			33.57	-5.95
KOR			25.09	2.39								
LIE	17.7	-2.28	41.9	-3.54	21.99	-4.6	47.71	-6.01	18.66	-4.56	47.07	-9.45
LVA	6.5	1.68										
MEX	19	-2.01	35.02	-3.15	17.47	-2.15	30.74	-3.54	16.51	-2.76	27.85	-4.46
NLD					17.89	-1.96						

NZL							43.17	3.4				
PRT							27.84	2.66				
QAT			8.89	5.23							16.06	6.4
ROM	7.15	1.96	27.09	3.15								
RUS			30.64	3.01			30.48	3.48	7.11	3.57	31.05	6.57
SVK									17.21	3.49		
SVN											48.35	-5.06
THA									15.76	-3.96		
TUR			45.82	-3.76	23.34	-2.75	49.79	-5.16	17.66	-2.77	38.91	-6.64
TWN					8.51	2.71	41.54	4.45				
URY	10.15	2.2			15.79	2.37						
USA			50.39	-3.56							51.05	-6.9

The significance level was set as $p < 0.05$. Blank cells indicate that β_1 was not significantly different from 0 (i.e., the Strength or the Slope of the socio-economic gradient did not experience detectable changes over time), and therefore information about β_0 was superfluous.

Table 4.G: Multilevel models mapping the evolution over time of the Strength of the socio-economic gradient for Reading, Mathematics and Science

	Reading			**Mathematics**			**Science**		
	Estimate	*Error*	*DF*	*Estimate*	*Error*	*DF*	*Estimate*	*Error*	*DF*
γ_{00} *(intercept)*	10.42***	(1.11)	188	12.50***	(1.10)	148	11.81***	(1.06)	108
γ_{10} *(t)*	0.29	(0.24)	188	0.07	(0.31)	148	-0.24	(0.35)	108
γ_{01} *(OECD)*	3.63**	(1.38)	54	3.37**	(1.38)	53	2.20	(1.34)	53
γ_{11} *(t*OECD)*	-0.60**	(0.29)	188	-0.62	(0.38)	148	0.24	(0.45)	108
σ^2_{u0}	19.40	(12–31)		20.63	(13–32)		20.90	(14–32)	
σ^2_{u1}	0.34	(0.10–1.14)		0.82	(0.34–2.01)		1.08	(0.38–3.08)	
σ_{u01}	-0.768			-1.397			-1.023		
σ^2_e	5.37	(4.25–6.79)		3.65	(2.75–4.85)		3.04	(2.09–4.41)	

***p <0.01, **p <0.05, *p <0.1

Table 4.H: Multilevel models mapping the evolution over time of the Slope of the socio-economic gradient for Reading, Mathematics and Science

	Reading			**Mathematics**			**Science**		
	Estimate	*Error*	*DF*	*Estimate*	*Error*	*DF*	*Estimate*	*Error*	*DF*
γ_{00} *(intercept)*	31.92***	(2.12)	188	33.54***	(1.83)	148	31.64***	(1.98)	108
γ_{10} *(t)*	-0.32	(0.48)	188	-0.73	(0.52)	148	-1.49**	(0.63)	108
γ_{01} *(OECD)*	7.31***	(2.66)	54	6.34***	(2.30)	53	10.86***	(2.51)	53
γ_{11} *(t*OECD)*	0.27	(0.60)	188	0.98	(0.64)	148	-0.14	(0.81)	108
σ^2_{u0}	79.88	(51–123)		58.55	(38–91)		74.57	(48–130)	
σ^2_{u1}	2.89	(1.57–5.31)		2.90	(1.42–5.93)		3.92	(1.56–9.89)	
σ_{u01}	-7.246			-2.347			-3.736		
σ^2_e	10.85	(8.53–13.79)		8.81	(6.63–11.70)		9.02	(6.21–13.11)	

***p <0.01, **p <0.05, *p <0.1

Table 4.I: Between-country variance of the Strength and the Slope of the socio-economic gradient as a function of time

Strength	**2000**	**2003**	**2006**	**2009**	**2012**
Reading	19.4	18.2	17.68	17.84	18.68

Mathematics		20.63	18.65	18.31	19.61
Science			20.9	19.94	21.14
Slope	**2000**	**2003**	**2006**	**2009**	**2012**
Reading	79.88	68.27	62.44	62.39	68.12
Mathematics		58.55	56.75	60.75	70.55
Science			74.57	71.01	75.29

Modelled using information from Table 4.G and Table 4.H above.

Table 4.J: A multilevel model on the relationship between mean and change in the Strength of the socio-economic gradient and PISA scores

	Reading			**Mathematics**			**Science**		
	Estimate	*Error*	*DF*	*Estimate*	*Error*	*DF*	*Estimate*	*Error*	*DF*
(Intercept)	420.44***	(17.50)	167	429.82***	(20.79)	140	443.28***	(17.57)	109
Δ Strength	0.26	(0.35)	167	-0.11	(0.31)	140	-0.09	(0.30)	109
Time	6.64***	(1.39)	167	5.24***	(1.57)	140	4.31***	(1.31)	109
OECD	78.67***	(13.57)	62	64.45***	(15.35)	61	60.35***	(12.55)	61
AVG Strength	-0.63	(1.29)	62	0.09	(1.39)	61	-0.34	(1.29)	61
*Time*OECD*	-5.57***	(1.67)	167	-5.66***	(1.98)	140	-2.88*	(1.69)	109
σ^2_{u0}	2570.98	(1723–3834)		3470.50	(2395–5027)		2341.61	(1632–3359)	
σ^2_{u1}	14.06	(4.73–41.82)		38.77	(20.80–72.27)		26.23	(14.63–47.01)	
σ_{u01}	-124.820			-195.651			-57.625		
σ^2_e	102.32	(78.19–133.91)		41.29	(29.30–58.18)		24.69	(16.92–36.04)	

***p <0.01, **p <0.05, *p <0.1

Table 4.K: Multilevel models mapping the linear and quadratic evolution over time of the gross enrolment ratio between 1998 and 2013

	Linear model			**Quadratic model**		
	Estimate	*Error*	*DF*	*Estimate*	*Error*	*DF*
(intercept)	38.40***	(4.72)	1271	41.68***	(4.75)	1270
Time	1.21***	(0.19)	1271	0.50**	(0.21)	1270
OECD	32.07***	(6.30)	59	31.53***	(6.32)	59
*Time*OECD*	-0.32	(0.25)	1271	-0.30	(0.26)	1270
Time^2				0.03***	(0.004)	1270
σ^2_{u0}	583.66	(404–841)		586.47	(408–842)	
σ^2_{u1}	0.90	(0.62–1.33)		0.94	(0.67–1.31)	
σ_{u01}	-12.80			-13.00		
σ^2_e	41.53	(38.34–44.98)		39.49	(36.46–42.77)	
Log L	-4636.865			-4607.665		
(df), p diff					(1), <0.0001	

***p <0.01, **p <0.05, *p <0.1

Table 4.L: A multilevel model of the relationship between GER and PISA in a subsample of lower-income countries.

	Reading			**Mathematics**			**Science**		
	Estimate	*Error*	*DF*	*Estimate*	*Error*	*DF*	*Estimate*	*Error*	*DF*
(intercept)	360.77***	(17.82)	83	369.99***	(17.96)	67	400.06***	(22.29)	46
Δ GER	0.33*	(0.18)	83	-0.03	(0.22)	67	0.10	(0.24)	46
Time	3.41***	(1.19)	83	4.06**	(1.72)	67	2.57	(1.94)	46
AVG GER	1.20***	(0.32)	30	1.06***	(0.31)	30	0.91**	(0.37)	23
σ^2_{u0}	1681.42	(937–3016)		2401.18	(1401–4114)		1666.64	(927–2993)	
σ^2_{u1}	14.72	(3.99–54.34)		37.89	(12.68–113.2)		41.16	(18.95–89.42)	
σ_{u01}	-41.827			-210.138			-44.804		

σ^2_e	102.49	(70.81–148.35)	58.17	(36.41–92.93)	23.51	(13.27–41.68)

***p <0.01, **p <0.05, *p <0.1

Given that most countries in the sample were not OECD members, the OECD variable was excluded from the model.

Table 4.M: A multilevel model of the relationship between GER and PISA in a subsample of high-income countries.

	Reading			**Mathematics**			**Science**		
	Estimate	*Error*	*DF*	*Estimate*	*Error*	*DF*	*Estimate*	*Error*	*DF*
(intercept)	473.74***	(21.76)	87	472.09***	(24.77)	67	467.63***	(24.11)	45
Δ GER	0.38*	(0.18)	87	0.03	(0.17)	67	0.16	(0.18)	45
Time	0.42	(1.50)	87	0.22	(1.67)	67	1.62	(1.38)	45
AVG GER	0.24	(0.25)	26	0.32	(0.29)	26	0.40	(0.29)	26
σ^2_{u0}	2563.66	(1347–4878)		2533.54	(1413–4542)		1506.79	(867–2620)	
σ^2_{u1}	42.69	(17.18–106.09)		58.71	(29.18–118.1)		26.60	(8.54–82.85)	
σ_{u01}	-304.144			-321.539			-150.453		
σ^2_e	78.82	(54.73–113.51)		31.73	(20.31–49.56)		29.47	(14.51–59.86)	

***p <0.01, **p <0.05, *p <0.1

Given that most countries in the sample were OECD members, the OECD variable was excluded from the model.

Table 4.N: A breakdown of country interventions

Country / Variable	**T 1**	**T 2**	**E 1**	**E 2**	**E 3**	**E 4**	**E 5**	**Q 1**	**Q 2**	**Q 3**	**Q 4**	**P 1**	**P 2**	**P 3**	**Country total**
Belgium Fr	0	0	0	0	0	1	0	1	1	0	0	1	1	0	5
Canada	0	0	1	1	0	0	0	0	0	0	0	1	1	0	4
Chile	1	1	1	1	0	0	1	1	1	1	0	1	0	1	10
Denmark	0	1	1	0	0	0	1	1	0	1	0	1	0	0	6
Finland	1	0	1	1	0	0	1	0	1	0	1	1	0	1	8
Germany	0	1	1	1	1	0	0	1	0	0	1	1	1	1	9
Indonesia	1	0	0	1	0	0	1	1	0	1	0	0	1	0	6
Israel	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1
Jordan	1	0	0	0	1	0	0	1	0	0	0	0	0	1	4
Korea	1	1	1	1	1	0	1	1	1	1	1	1	0	1	12
Latvia	1	0	0	0	0	0	0	0	0	1	1	1	1	1	6
Netherlands	0	0	0	1	0	0	0	1	1	1	0	0	1	0	5
Portugal	1	1	0	0	0	0	0	1	0	1	1	0	1	1	7
Russian Federation	1	0	1	1	1	0	1	1	1	0	1	0	1	1	10
Russian Federation	1	0	1	0	0	0	0	1	0	0	0	0	1	0	4
Thailand	1	0	0	0	0	0	1	1	0	0	0	1	0	0	4
Variable total	10	5	8	8	4	1	7	12	6	7	6	9	10	8	

Variable	Recommendation
T1	Increase overall time spent in education
T2	Increase instruction time
E1	Reduce disabled or immigrant students' segregation
E2	Reduce the gender gap
E3	Reduce the streaming / grouping / stratification of students
E4	Reduce school competition
E5	Help socio-economically deprived students
Q1	Strengthen the teaching profession
Q2	Allocate resources more efficiently or in a more equitable way
Q3	Pair more autonomy with accountability measures
Q4	Promote cooperation between parents, teachers, principals and schools
P1	Help low-performing students
P2	Standardise the educational outcomes
P3	Review curricular priorities

Table 4.O: Country rankings according to their score on the PAIQ and according to the EPA Index

Ranking	By PAIQ score	By EPA Index	Change PAIQ → EPA
1	Korea	Finland	+4
2	Chile	Korea	-1
3	Russian Federation	Denmark	+4
4	Germany	Chile	-2
5	Finland	Germany	-1
6	Portugal	Netherlands	+5
7	Denmark	Russian Federation	-4
8	Indonesia	Canada	+4
9	Latvia	Belgium Fr	+1
10	Belgium Fr	Portugal	-4
11	Netherlands	Latvia	-2
12	Canada	Israel	+3
13	Jordan	Indonesia	-5
14	Thailand	Thailand	0
15	Israel	Jordan	-2

Table 4.P: Common (link) items between comparable assessments

Reading	2003	2006	2009	2012
2000	28	28	26	3
2003	-	28	26	3
2006		-	26	3
2009			-	44
Mathematics	**2003**	**2006**	**2009**	**2012**
2003	-	48	35	84
2006		-	35	48
2009			-	35
Science	**2003**	**2006**	**2009**	**2012**
2006	-	-	53	53
2009				53

Sources: PISA technical reports 2006–2009 (OECD, 2009c, 2012e), and PISA international report 2012 (OECD, 2014b, Appendix A5)

Chapter 5

The tables below summarise findings from the national and international reports for Ireland, along with the differences between the two versions. Remarks are coded as follows:

+1 is a positive remark that suggests achievement in a 'positive' aspect. Example: 'at both grades close to half the students in Singapore (45% at the eighth grade and 44% at the seventh grade) reached the Top 10% level.'

0 is a descriptive or mixed statement. Example of a descriptive statement: 'Countries close to the trend line are where the predictor GDP per capita suggests that they would be; examples include…'. Example of a mixed statement: 'students in the top quarter of the index in […] Ireland report a much greater interest in reading than students in the top quarter of the index in Belgium […] In Ireland there is much wider variation in students' interest in reading than in Mexico'. In this case a positive aspect (high values at the top 25%) is followed by a negative (large spread).

−1 is a negative remark and it serves to flag up a potential problem, to suggest that there is an area open for review. Example: 'Fewer than 40% of the eighth-grade students in [...] Ireland [...] had teachers who felt it was very important to think creatively, and fewer than 40% [...] had teachers who felt it was very important to understand how mathematics is used in the real world.' The follow-up is: 'with the current calls from business and industry for helping students improve their ability to apply mathematics and solve practical problems in job-related situations, it might be rather surprising that...'

Table 5.A: Findings and differences between the international and national versions of the PISA 2000 report

		GENERAL		
Remark		**Area**	**Reference pages**	
OECD	*ERC*		*OECD*	*ERC*
0	-	GDP (close to the trend line)	92	-
-1	-	Engaging learning environment	108	-
		READING		
Remark		**Area**	**Reference pages**	
OECD	*ERC*		*OECD*	*ERC*
+1	+1	Overall achievement	44	vii
+1	-1	Gender differences in achievement	125	see txt
+1	-	Succeeding against the odds (counterbalancing mother's SES)	151	
+1	-1	Overall impact of SES	195	x,9,12
0	-1	Distribution of achievement (Finland < %below L1 < OECD average)	47	14
0	0	Interest	102	ix
0	-	Teacher shortage	172	
0	-1	Slope of SES-Score relationship	191	x,9,12

-1	-1	School type	63	ix
-1	-	Gender differences in engagement	131	
-1	-1	Single-parent family	152	
		MATHEMATICS AND SCIENCE		
Remark		**Area**	**Reference pages**	
OECD	*ERC*		*OECD*	*ERC*
-1	-1	Achievement compared to reading	90	viii

Sources: OECD (2001a); Shiel et al. (2001a; 2001b)

While the PISA report is rather positive with respect to gender differences ("Some countries do appear to provide a learning environment that benefits both genders equally", OECD, 2001a, p. 125), the ERC summary report is more critical: on the one hand, their interaction with other explanatory variables is stressed; on the other hand, where findings indicate no significant gender differences (e.g., in Science), the authors advise that 'the PISA findings should not induce complacency' (Shiel et al., 2001a, p. 19). Another area whose interpretation differs substantially from OECD's is the score distribution: OECD mentions that Ireland has a proportion of students below Level 1 greater than that of Finland, but nevertheless lower than the OECD average; however, Shiel et al. (2001a) add: 'While the proportion of Irish students achieving Level 1 or below on the PISA combined reading literacy scale (11.0%) was lower than the OECD country average (17.9%), it is nevertheless a cause for concern' (p. 14). Given that the Slope of the socio-economic gradient in Ireland is lower than in other countries, the PISA report treats the impact of an Irish student's socioeconomic status more leniently than the ERC does.

Additionally, the ERC remarks other issues such as the distribution of achievement in Mathematics: 'Irish students at the national 90th percentile achieved a score that is below the corresponding OECD country average. Ireland ranked 20th, indicating a relatively poor performance by higher-achieving students.' (Shiel et al., 2001a, p. viii). The ERC is more critical of the Science performance: 'While this relatively good performance is welcome, it cannot be concluded that it represents a level of achievement in science that is superior to that in earlier studies.' (*ibid*., p. 15). It is also noted that this round of PISA was covered aspects of biology more than chemistry or physics. Finally, the test-curriculum alignment is also discussed, and the overall conclusion is that while the Mathematics and Science taught in the Junior Certificate syllabus are different from the domains assessed by PISA, 'there is considerable overlap' between PISA and the Junior Certificate examinations (*ibid*., p. 13).

Table 5.B: Findings and differences between the international and national versions of the PISA 2003 report

		GENERAL		
Remark		**Area**	**Reference pages**	
OECD	*ERC*		*OECD*	*ERC*

0	-	Determining course content	235–238	
0	-	Parental influence	238	
0	-	Instruction time	241	
0	-	Principal inspection	247	
-1	-	Tracking	261	
		READING		
Remark		**Area**	**Reference pages**	
OECD	*ERC*		*OECD*	*ERC*
-1	+1	Performance	284	42
		MATHEMATICS		
Remark		**Area**	**Reference pages**	
OECD	*ERC*		*OECD*	*ERC*
+1	0	Classroom disciplinary climate	215	44
+1	0	Between-school variance	163	41–42
+1	-	Gender (interest)	121	
0	-1	Distribution of achievement	95	40–41
0	-	Change in Space and Shape scale since 2000	62	
-1	-1	Space and Shape scale	89–90	40
-1	0	Gender (performance)	98	41
-1	-	Gender (sense of belonging to school)	128	
-1	-1	Single-parent family	167	44
		SCIENCE		
Remark		**Area**	**Reference pages**	
OECD	*ERC*		*OECD*	*ERC*
+1	0	Performance	293	43
0	-	Change in achievement since 2000	296	

Sources: OECD (2004b); Cosgrove et al. (2004)

Some differences between the international and the ERC reports are explainable by the additional analyses carried out by ERC. The two reports agree on the negative Mathematics performance, especially in the Space and Shape subdomain, but the ERC also stresses the high number of students achieving at the lowest levels and the relative underperformance of those at the highest ones.

The ERC tends to downplay some aspects that were positively commented on by OECD, like the generally positive classroom climate and the low between-school variance: while the international reports concludes that 'parents […] can be confident of high and consistent performance standards across schools in the entire education system' (OECD, 2004b, p. 163), the ERC simply attributes this 'to relatively homogeneous composition of the school-going population in Ireland' (Cosgrove, Shiel, Sofroniou, et al., 2004, p. 42). On the other hand, PISA findings of a higher sense of belonging in school for girls are not reported.

Table 5.C: Findings and differences between the international and national versions of the PISA 2006 report

	GENERAL	

Remark		Area	Reference pages	
OECD	*ERC*		*OECD*	*ERC*
0	-1	Examination board influence on assessment practices	251	36
-1	-	Parental pressure	233	
		READING		
Remark		**Area**	**Reference pages**	
OECD	*ERC*		*OECD*	*ERC*
+1	+1	Achievement	299	33
		MATHEMATICS		
Remark		**Area**	**Reference pages**	
OECD	*ERC*		*OECD*	*ERC*
+1		Distribution of achievement	319	
0	0	Achievement	315	33
		SCIENCE		
Remark		**Area**	**Reference pages**	
OECD	*ERC*		*OECD*	*ERC*
+1	+1	Between-school variance	173	38
+1	-	Migrant status (performance)	175	
0	0	Motivation, self-concept, attitude (see text)	127–163	36
0	0	Ability grouping	223–225	38
0	-	Private institutions	231	

Sources: OECD (2007b); Eivers, Shiel, & Cunningham (2007, 2008)

According to the ERC, 'Ireland's students performed very well on the reading assessment, reasonably well on science, and about average on mathematics' (Eivers et al., 2007, p. 33), and the narrow is distribution in Science was also positively remarked (*ibid*.). As in 2003, Ireland was elected by the OECD an exemplary country for its lower between-school variance (2007b, p. 173), but the ERC focused instead on the negative effects of ability grouping (Eivers et al., 2007, p. 38). Other school issues raised by the ERC were student absenteeism and bullying.

At the family level, the ERC reported lack of change in home educational resources (measured in number of books at home), which was 'unfortunate' given its positive relationship with achievement, especially in the context of an increase of indicators (such as measures of affluence) negatively related with performance (Eivers et al., 2007, p. 37).

From the OECD reports, a tripartite positive relationship emerged between: student SES; their motivation, self-concept and attitudes; and Science performance. Moreover, the majority of students valued science but did not enjoy studying it and 'did not agree with the wider conception of it bringing economic and social benefits' (OECD, 2007b, p. 128). Science was more highly valued by immigrant students, whereas lack of enjoyment correlated with socio-economic disadvantage. These findings probably warranted further investigation, yet the ERC

focused only on differences in the interest towards biology, physics and chemistry. The absence of comments on the good performance of students with an immigrant background and their valuing science is noteworthy. Another puzzling exclusion is the high percentage of principals reporting high levels of parental pressure, since many governmental interventions at the time mentioned the importance of better communication between families and schools.

Table 5.D: Findings and differences between the international and national versions of the PISA 2009 report

		GENERAL		
Remark		**Area**	**Reference pages**	
OECD	*ERC*		*OECD*	*ERC*
+1	-	Student transfer policies	4.66	
0	-	Private institutions	4.43	
0	-	Instruction time	4.79–80	
0	-	Class size	4.82	
0	-	Resource allocation	4.84	
0	-	Parental pressure	4.98	
0	-	Leadership	4.102	
-1	-	Change in disciplinary classroom climate	5.100	
		READING		
Remark		**Area**	**Reference pages**	
OECD	*ERC*		*OECD*	*ERC*
+1	-	Between-school variance	2.84–86	
+1	-	School location (urban or rural)	2.14	
+1	-	Reflect and Evaluate subscale	1.71	
+1	+1	Understand and Remember	3.77–79	ix
0	-	Achievement	1.55	
0	-	Continuous texts	1.85	
0	0	Socio-economic status and performance	2.31, 2.57	ix
0	0	Migrant status	2.71–72, 2.75, 2.79	ix
-1	-1	Single-parent family	2.14	26
-1	-	Pre-primary education (more than one year)	2.96, 4.81	
-1	-1	Motivation (reading for enjoyment)	3.34, 3.66	ix
-1	-1	Change in motivation (reading for enjoyment)	5.90, 5.97	ix
-1	-1	Change since 2000	5.13	vii
-1	-1	Distribution of achievement	5.13, 5.43	viii
-1	-1	Gender differences in change in achievement	5.48	vii–viii
-1	-1	Socio-economic status role in explaining change	5.50	
		MATHEMATICS		
Remark		**Area**	**Reference pages**	
OECD	*ERC*		*OECD*	*ERC*
+1	0	Gender differences	1.137	viii
-1	-1	Change since 2000	5.60	viii

-1	-1	Distribution since 2000	5.63	viii
		SCIENCE		
Remark		**Area**	**Reference pages**	
OECD	*ERC*		*OECD*	*ERC*
+1	0	Achievement	1.14	viii–ix

Sources: OECD (2010b, 2010c, 2010d, 2010e, 2010f, 2010g); Perkins et al. (2010)

The drop in student achievement in both Reading and Mathematics is deemed to be 'notable' by OECD, and mentions of it can be found in the Executive Summary for volume V (OECD, 2010d). This is aggravated by the fact that 'student performance would have declined more rapidly than actually observed if the changes in the socio-economic composition of the student population had been accounted for' (*ibid*., p. 50), as it shows that there is a strong relationship between high socio-economic status and high achievement. As it happened in previous years, the ERC picks up on the same negative aspects identified by OECD whilst also minimising the positive ones.

The relationship between migrant status and performance is, once again, multifaceted. Firstly, immigrant students have a similar socio-economic status to that of native students. Secondly, there are major achievement differences between first- and second-generation migrants, with the latter performing as well as native students: 'These large gaps […] possibly [highlight] the different backgrounds across immigrant cohorts […]. However, they could also signal positive educational and social mobility across generations' (OECD, 2010e, p. 72). Moreover, 'the relationship between attendance in pre-primary education and performance is greater for students with an immigrant background' (*ibid*., p. 98).

Another interesting finding is that socio-economically advantaged students are much more likely to read for enjoyment, which is in turn associated to higher achievement. This mirrors the finding from Science 2006. Moreover, the percentage of students reading for enjoyment 'decreased by almost 18 percentage points among disadvantaged students' (OECD, 2010d, p. 97).

Parental pressure regarding academic standards is a difficult topic to categorise, because it seems that OECD itself has yet to decide whether it is a positive or a negative feature of a system. For PISA 2006, OECD writes that 'parental pressure on schools is largely absent […in] Finland – the best performing country' (OECD, 2007c), which seems to hint as lack of pressure as a positive thing. However, for PISA 2009, there is an opposite take on the matter: 'Research suggests that students perform better when parents, teachers and schools have high expectations for them. A driving force behind school expectations is parental pressure' (OECD, 2010f, p. 98). In either case, parental pressure is high in Ireland.

Chapter 6

The tables below summarise findings from the national and international reports for the CFB, along with the differences between the two versions. Note, however, that in the case of the CFB this exercise is less informative than in the case of Ireland because findings from the three communities are aggregated in international reports.

Remarks are coded as follows:

+1 is a positive remark that suggests achievement in a 'positive' aspect. Example: 'at both grades close to half the students in Singapore (45% at the eighth grade and 44% at the seventh grade) reached the Top 10% level.'

0 is a descriptive or mixed statement. Example of a descriptive statement: 'Countries close to the trend line are where the predictor GDP per capita suggests that they would be; examples include…'. Example of a mixed statement: 'students in the top quarter of the index in […] Ireland report a much greater interest in reading than students in the top quarter of the index in Belgium […] In Ireland there is much wider variation in students' interest in reading than in Mexico'. In this case a positive aspect (high values at the top 25%) is followed by a negative (large spread).

−1 is a negative remark and it serves to flag up a potential problem, to suggest that there is an area open for review. Example: 'Fewer than 40% of the eighth-grade students in [...] Ireland [...] had teachers who felt it was very important to think creatively, and fewer than 40% [...] had teachers who felt it was very important to understand how mathematics is used in the real world.' The follow-up is: 'with the current calls from business and industry for helping students improve their ability to apply mathematics and solve practical problems in job-related situations, it might be rather surprising that...'

Table 6.A: Findings and differences between the international and national versions of the PISA 2000 report for the CFB

		GENERAL	
Remark		**Area**	**Reference pages**
OECD	*ASPE*		*OECD* *ASPE*
-1	-1	high proportion of low achievers	44 49
+1		mean performance	52
-1		like school	108
0		use of elaboration strategies (Fl)	112
0		teacher support	162
0		school climate importance	167
0		teacher morale importance	169
-1		homework	171
-1		teacher shortage	172
+1		infrastructures	173
0		autonomy	177, 178

-1		specialist teachers	211	
		READING		
Remark		**Area**	**Reference pages**	
OECD	*ASPE*		*OECD*	*ASPE*
	-1	Retention		12
	-1	Gender differences		71
-1	-1	Retrieving information stronger	54	54
-1	-1	country-level variance, widest spread, note on Fr. Vs Fl.	56, 57	55
-1	-1	most variance btw schools	60	Chap 4
0		close to GDP prediction	92	
-1	-1	interest (Fl.)	100	98
-1	-1	engagement	104, 106	109, 110
	-1	Variety of texts and performance		105
-1 /+1	-1	Parental occupation vs family wealth	139, 143	86
-1		Mother education	149	
-1		Single-parent families	152	
-1	+1	first-generation and immigrant students	155	80
		MATHEMATICS		
Remark		**Area**	**Reference pages**	
OECD	*ASPE*		*OECD*	*ASPE*
+1	0	mean	80, 125	59
-1	-1	distribution, similar pattern to reading	80, 82	60
+1		gender differences	125	
		SCIENCE		
Remark		**Area**	**Reference pages**	
OECD	*ASPE*		*OECD*	*ASPE*
0/-1	-1	performance, AVG but worse than reading	90	63

Sources: OECD (2001a); Lafontaine (2001a); Lafontaine et al. (2003)

Table 6.B: Findings and differences between the international and national versions of the PISA 2003 report for the CFB

		GENERAL		
Remark		**Area**	**Reference pages**	
OECD	*ASPE*		*OECD*	*ASPE*
0		Use of control strategies	148	
0		School climate	218	
0		Assessment for school accountability	232	
0		School autonomy over teachers and budgets	234,235	
0		Monitoring by principals		247
0/-1		Teacher commitment and student results	225	
-1		Sense of belonging in schools, and higher in females	128	
-1		Use of standardised assessments	230	
-1		Pre-primary education attendance gap	243	
-1		Teacher shortage	245	
-1		School resources	255	
		READING		
Remark		**Area**	**Reference pages**	

OECD	*ASPE*		*OECD*	*ASPE*
+1	+1	Percentage at level 5	277	
+1	0	Mean performance	280	4
-1	-1	Spread	278	
		MATHEMATICS		
Remark		**Area**	**Reference pages**	
OECD	*ASPE*		*OECD*	*ASPE*
+1		Percentage over level 6	51	
+1		Mean performance increase change/relationship	72	
+1		Gender gap	96, 98	
+1		Spending per student	103	
+1/0		Btw-school variance decreased (maybe)	164 (204)	
+1/0	-1	Mean performance increase space/shape 28 points / driven by higher-ability students	61, 63	4
-1		Uncertainty scale lower than others	89	
-1	-1	Subnational differences: French Community average	93	4
-1	-1	Spread	95	
-1		Intrinsic motivation	118	
-1		Instrumental motivation	123	
-1		Between-school variance	161, 163	
-1		Parental occupation	165	
-1		Single-parent families	167	
-1		Immigrant students (greatest gap), even after accounting for SES, and language	168, 172, 173	
		SCIENCE		
Remark		**Area**	**Reference pages**	
OECD	*ASPE*		*OECD*	*ASPE*
+1	-1	Mean performance	293	4
+1/0	+1	Mean performance increase / driven by higher-ability students	295	8

Sources: OECD (2004b); Baye et al. (2004b)

The full national report was not available at the ASPE, therefore the comparison is between the international report and the summary report.

The difference in the remarks concerning mean performance are due to the fact that in the international report this is pushed up by the performance of the Flemish Community. It is interesting to note, nevertheless, that in the national report the ASPE underlines the greater distance of the CFB from the OECD average in Mathematics and Science. This contrasts with the accurate explanation about confidence intervals and statistically significant differences that follows and makes one wonder why opening with a statement apparently showing the critical status of mathematics and science education if caution is then advocated.

Performance differences due to the segregating effects of streaming and retention are highlighted in both reports, but the ASPE does not comment (at least in the summary report) about a positive signal, such as a possible decrease observed in between-school variance.

Instead, a non-significant increase in Science was an opportunity to remark that the increase was driven by mid-low performers.

Table 6.C: Findings and differences between the international and national versions of the PISA 2006 report for the CFB

		GENERAL		
Remark		**Area**	**Reference pages**	
OECD	*ASPE*		*OECD*	*ASPE*
+1		Average number of students per teacher	256	
+1		Importance of catchment area	217	
+1		Ability grouping	223	
0		National income	59	
0		School autonomy	251	
0		Parent / public reporting of results	242	
0		Result-based teacher accountability	240	
-1		Streaming	220	
-1	-1	Teacher shortage	254	216, 217
	-1	Retention		15
		READING		
Remark		**Area**	**Reference pages**	
OECD	*ASPE*		*OECD*	*ASPE*
+1	0	Percentage at level 5	294	75
+1	-1	Mean performance	299	48
-1	-1	Spread	299	75
-1	-1	Gender differences	303	81
		MATHEMATICS		
Remark		**Area**	**Reference pages**	
OECD	*ASPE*		*OECD*	*ASPE*
+1	0	Percentage over level 6	313	72
+1	0	Mean performance	315	48
-1	-1	Spread	319	72
-1		Change 2003–2006 among higher achievers	320	
		SCIENCE		
Remark		**Area**	**Reference pages**	
OECD	*ASPE*		*OECD*	*ASPE*
+1	-1	Future-oriented motivation, expectation of science-related career	150, 152	180
+1	+1	Decrease in between-school variation	173	134
	+1	Intrinsic motivation		173
	+1	Engagement in science-related activities		185
	0	Percentage high achievers		63
0	0	Knowledge about science higher than knowledge of science	71	229
0		Learning time	269	
-1	-1	Difference mean performance Flanders vs French Community	55	58
-1		Gender differences within schools	62	
-1	-1	Advances in science bring social benefits	128	162
-1	0	Higher interest in science for higher-SES students	140	164
-1	-1	Awareness of environmental issues is related to SES	155	194, 195

-1	-1	Between-school variance	172	134
-1	-1	Immigrant students' performance, segregation and school resources	175–180	91
-1	-1	Relationship SES-achievement	185	96
	-1	Number of students enrolled in scientific tracks		16
	-1	Explain Phenomena Scientifically scale		52
	-1	Earth & Space worse, but also physics and biology are low		55
	-1	Spread		56
	-1	Percentage low achievers		62
	-1	Guidance and satisfaction with their classes		205, 206
	-1	Constructivist approach, opportunities to experiment, realistic contexts		210, 212
	-1	Promotion of science within schools		214

Sources: OECD (2007b) ; Baye et al. (2009)

The different sign in the mean performances are due to the fact that the national report refers to the French Community only.

Table 6.D: Findings and differences between the international and national versions of the PISA 2009 report for the CFB

		GENERAL		
Remark		**Area**	**Reference pages**	
OECD	*ASPE*		*OECD*	*ASPE*
+1		Student-teacher ratio positively related to school SES	2.43	
+1		Participation to pre-primary education	2.96	
+1		School autonomy on curriculum and assessment	4.41	
+1		School climate	4.94, 4.96	
+1/-1		School choice	4.72, 4.74	
0		Class size	4.82	
-1		Grade repetition	4.64	
-1		Transfer policies	4.66	
		READING		
Remark		**Area**	**Reference pages**	
OECD	*ASPE*		*OECD*	*ASPE*
+1		Mean performance (combined)	1.13	
	+1	2000–2009 increase in mean performance		8
+1		Integrate and interpret	1.63	
+1		Percentage of top performers	1.156	
+1		No difference urban-rural	2.49	
+1		Awareness of strategies to summarise information	3.79	
	+1	2000–2009 gender gap narrowed		11
+1	+1	2000–2009 immigrant gap narrowed	5.14	11
+1	+1	2000–2009 variance, decrease in below level 2	5.74	9
+1		2000–2009 relationship SES-achievement between schools	5.80	
0		Relationship reading strategies-scores	3.13, 3.47	
0		Relationship enjoyment-scores	3.30	
0		Reading for enjoyment > time spent	3.34	
-1	-1	Spread	1.53	12
-1		Gender gap on non-continuous texts	1.85	

-1	-1	Relationship SES-achievement	2.14, 2.45	13
-1		Single-parent family	2.14	
	-1	Grade repetition gap		15
-1	-1	Immigrant students, school resources	2.70, 2.79	15
-1	-1	Between-school variance, inclusion	2.84, 2.86	16
-1		Percentage of students reading for enjoyment, relationship with SES	3.34, 3.66	
-1	-1	Streaming	4.36	15
-1		Teachers recommend books	4.94	
-1		2000–2009 gap from speaking the test language at home	5.82	
-1	+1	2000–2009 no change boys reading for enjoyment	5.89	11, 12
		MATHEMATICS		
Remark		**Area**	**Reference pages**	
OECD	*ASPE*		*OECD*	*ASPE*
+1		Mean performance (combined)	1.14	
+1		Percentage high performers	1.130, 1.132	
-1		Spread	1.53	
-1		Gender differences	1.137	
-1	0	2003–2009 performance	5.14	8
-1		2003–2009 increase in below level 2, decrease in top performers	5.63	
		SCIENCE		
Remark		**Area**	**Reference pages**	
OECD	*ASPE*		*OECD*	*ASPE*
+1		Mean performance (combined)	1.14	
	0	2006–2009 performance		8

Sources: OECD (2010b, 2010c, 2010d, 2010e, 2010f, 2010g); Baye et al. (2010)

The full national report was not available. The OECD notes that Belgium allocates more teachers into disadvantaged schools ‘presumably with the objective of moderating that disadvantage’ (OECD, 2010e, p. 43). This policy, which is in line with the EPA, was not commented upon by the ASPE. In spite of the already active policies on positive discriminations, the OECD generically reminds that ‘policies that target socio-economically disadvantaged schools are more likely to succeed in improving performance and equity’ (OECD, 2010e, p. 113). This suggest that the EPA is not responsive to country efforts.

The EPA concerning school choice is unclear. On the one hand, the OECD writes that the most successful countries ‘tend to grant higher levels of autonomy to schools in formulating and using curricula and assessments and lower levels of school competition’ (OECD, 2010f, p. 74). On the other hand, the positive effects of school choice were mentioned three pages earlier.

Both the ASPE and the OECD agree that 'higher levels of engagement in reading activities and positive approaches to learning' may benefit disadvantaged students in particular (OECD, 2010c, p. 93).

Table 6.E: A comparison between PISA literacies and the French, mathematics and science literacy according to the Core Skills

	OECD	**CFB**
R	Reading literacy is understanding, using, reflecting on and engaging with written texts, in order to achieve one's goals, develop one's knowledge and potential, and participate in society.	The French language is the first key that is offered to the child and adolescent to access all areas of learning. […] pupils will be prompted to perform a set of interactive skills, mental processes, learning strategies and relational attitudes, that can be used directly in the construction of their knowledge, but more importantly, once their schooling is completed, as the foundations of their life-long learning processes.
M	Mathematical literacy is an individual's capacity to formulate, employ, and interpret mathematics in a variety of contexts. It includes reasoning mathematically and using mathematical concepts, procedures, facts and tools to describe, explain and predict phenomena. It assists individuals to recognise the role that mathematics plays in the world and to make the well-founded judgments and decisions needed by constructive, engaged and reflective citizens.	Mathematics is taught on the basis of objects and situations experienced and observed in real life to raise questions about mathematical facts. The mathematics course is not just about imparting knowledge. From primary school to the end of the first stage of secondary school, stimulating the imagination, provoking thought and developing critical thinking about these observations lead pupils to understand their environment and have an effect on it. […] *Through problem-solving, pupils develop mathematical abilities, acquire in-depth knowledge and forge a confident and active personality*.
S	For the purposes of PISA, scientific literacy refers to an individual's: ▪ Scientific knowledge and use of that knowledge to identify questions, acquire new knowledge, explain scientific phenomena and draw evidence-based conclusions about science-related issues. ▪ Understanding of the characteristic features of science as a form of human knowledge and enquiry. ▪ Awareness of how science and technology shape our material, intellectual and cultural environments. ▪ Willingness to engage in science-related issues, and with the ideas of science, as a reflective citizen.	Teaching science is both about the development of specific and transversal skills and the acquisition of knowledge and offers the most appropriate methodologies to encourage young people to get a lasting grasp of them. The progressive construction of knowledge and skills is the founding element (paradigm) of any scientific approach. This, in effect, allows pupils, regardless of their age and level of study, to be the main actors in their learning process starting out from situations that encourage them to get involved in research. Furthermore, the study of science offers some specificity because it opens the eyes of young people to their natural environment and puts them in direct contact with real objects, natural phenomena and living beings. In the virtual age with its packaged products, this is an important input that should be highlighted.

Sources: OECD (2013c); Ministère de la Communauté française (2014)

Appendix 4: Sample participant information sheet and consent form (Belgium)

[university logo]
Date: _____

Participant Information Sheet

Title: The use and usefulness of international large-scale assessments (PISA, PIRLS, TIMSS) for national policy-making in the French Community of Belgium (1995-2009)

You are invited to take part in a research study on the relationship between international policy advice and national education reforms in different countries. Please read this form carefully and ask any questions you may have before agreeing to be in the study.

The study is conducted by Cesare Aloisi as part of his doctoral studies at Durham University. This research project is supervised by Prof Peter Tymms (peter.tymms@cem.dur.ac.uk) and Dr Christine Merrell (christine.merrel@cem.dur.ac.uk) from the School of Education at Durham University.

The purpose of this study is to understand in details how international large-scale assessment reports were received in the French Community of Belgium, to what extent they may have informed national reforms, interventions and policies, what the impact of these policies was, and in general whether international reports have been useful agents for educational improvement.

If you agree to be in this study, you will be asked to take part in an interview in which I will ask your views on these issues. The whole interview should last one hour.
You are free to decide whether or not to participate. If you decide to participate, you are free to withdraw at any time without any negative consequences for you.

All responses you give or other data collected will be kept confidential. The records of this study will be kept secure and private. All files containing any information you give are password protected. In any research report that may be published, no information will be included that will make it possible to identify you individually. There will be no way to connect your name to your responses at any time during or after the study.

If you have any questions, requests or concerns regarding this research, please contact me via email at: cesare.aloisi@durham.ac.uk; or by telephone at 0489 70 02 38.

This study has been reviewed and approved by the School of Education Ethics Sub-Committee at Durham University (date of approval: 18 March 2015)

Thank you for your time.

With very best wishes,

Cesare Aloisi

Leazes Road
Durham City, DH1 1TA
Telephone +44 (0)191 334 2000 Fax +44 (0)191 334 8311
www.durham.ac.uk
Durham University is the trading name of the University of Durham

Declaration of Informed Consent

I agree to participate in this study, the purpose of which is to understand the reception, use and usefulness of international large-scale assessment reports for policy-making in the French Community of Belgium.

- I have read the participant information sheet and understand the information provided.
- I have been informed that I may decline to answer any questions or withdraw from the study without penalty of any kind.
- I have been informed that all of my responses will be kept confidential and secure, and that I will not be identified in any report or other publication resulting from this research.
- I have been informed that the investigator will answer any questions regarding the study and its procedures. Cesare Aloisi, from the School of Education, Durham University, can be contacted via email: cesare.aloisi@durham.ac.uk or telephone: 0489 70 02 38.
- I will be provided with a copy of this form for my records.

Any concerns about this study should be addressed to the Ethics Sub-Committee of the School of Education, Durham University via email (Sheena Smith, School of Education, tel. +44 (0)191 334 8403, e-mail: Sheena.Smith@Durham.ac.uk).

Date
Participant Name (please print)
Participant Signature

I certify that I have presented the above information to the participant and secured his or her consent.

Date
Signature of Investigator

[French version would follow here]

Appendix 5: The multilevel modelling approach used in Chapter 4

Analytical methods

General approach

Figure 4.1 and Figure 4.2 (main document) map the trends in PISA scores in Reading and Mathematics for OECD and partner countries. To model these and the Science trends, multilevel/hierarchical modelling (Goldstein, 2011; Raudenbush & Bryk, 2002) was employed. This is a form of 'partial pooling' (Gelman & Hill, 2007) analysis that takes the grouping of the observations into account by assigning a variance component not only to the individual observations, but also to the different clusters, or levels, within which the observations are nested. To model these trends, one can write:

$$Scores_{ij} = \beta_{0j} + \beta_{1j} t_{ij} + e_{ij} \quad (1)$$

Where the subscript i indexes country-year observations (the Level-1 unit) and the subscript j indexes countries—the cluster in which the observation took place (the Level-2 unit); t is an ordered variable representing PISA administrations, and it takes the values 0 to 4 for Reading (0 corresponds to the year 2000), 0 to 3 for Mathematics (0 corresponds to 2003) and 0 to 2 for Science (0 corresponds to 2006). Note that both the intercept β_0 and the slope β_1 take the j-subscript, because they are both allowed to vary by group (country), and in fact, they can be decomposed thus:

$$\beta_{0j} = \gamma_{00} + u_{0j} \quad (2)$$

$$\beta_{1j} = \gamma_{10} + u_{1j} \quad (3)$$

In Equations (2) and (3) both the intercept and the slope coefficients are composed of a fixed part γ, which is common to all groups, plus a random part u that varies by group. Equations (1)–(3) can be expressed in long form whilst separating fixed from random (between brackets) components:

$$Scores_{ij} = \gamma_{00} + \gamma_{10} t_{ij} + (u_{0j} + u_{1j} t_{ij} + e_{ij}) \quad (4)$$

Equation (4) has six parameters that need estimating: γ_{00} and γ_{01}, the variances of u_0, u_1 and e, as well as the covariance between u_0 and u_1, where it is assumed that:

$$e_{ij} \sim N(0, \sigma_e^2) \quad (5)$$

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N(0, \Omega_u) : \ \Omega_u = \begin{bmatrix} \sigma_{u0}^2 & \\ \sigma_{u01} & \sigma_{u1}^2 \end{bmatrix} \quad (6)$$

Model-building followed the heuristic shown in Appendix 3, Table 4.C. Model A and B were respectively the null model and a model with one fixed slope for all countries. A simple

visual analysis of Figure 4.1 and Figure 4.2 may suggest that slopes vary by country, and indeed an analysis of variance confirmed that Model C, representing the Equation (4) introduced above, was more efficient (Appendix 3, Table 4.D).

Following diagnostics (Appendix 6), a model (D) was produced that accounted for different variances of the Level-1 residuals between OECD and partner countries, autocorrelation of the Level-1 residuals in Reading and a curvilinear structure in the Reading data[28] (Table i).

Table i: Quadratic (Reading) and linear (Mathematics) growth models with weighted variances and autocorrelated [MA(1)] residuals (Model D)

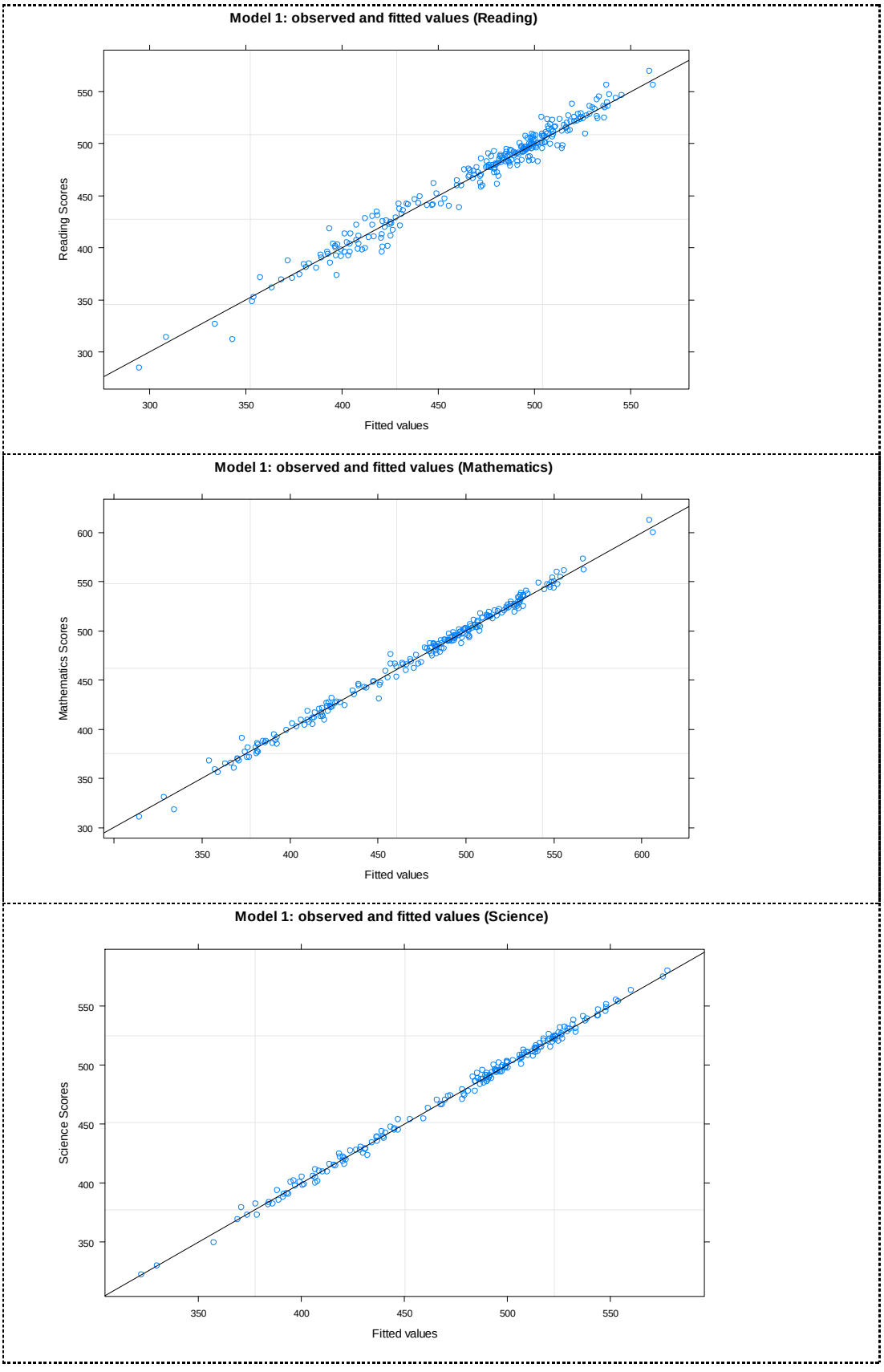
Model D	**READING**	**MATHEMATICS**
γ_{00}	462.71* (6.78)	463.47* (8.51)
γ_{10}	63.43* (16.40)	2.23* (0.90)
γ_{20}	37.02* (10.09)	
σ^2_{u0}	3857 (2654–5605)	4694 (3291–6697)
σ^2_{u1}	16.22 (7.00–37.54)	36.94 (21.79–62.62)
σ_{u01}	-210.52	-323.75
σ^2_e	191.07 (124.95–292.16)	74.37 (47.86–115.57)
Θ	0.25 (0.03–0.46)	

*p<0.01. All values estimated with restricted maximum likelihood (REML)

Compared to Model C (Appendix 3, Table 4.D), there is a small reduction in the Level-2 variance, counterbalanced by a larger increase in Level-1 variance. This means that while Model D gives a better account of the relative position of countries between each other, it is less successful in "tracking" the evolution of within-country score-change over time. Moreover, Model C is more efficient, as it has fewer degrees of freedom, and it can be equally applied to all three PISA domains while still providing a good fit (Figure i).

[28] A quadratic model of the form $Reading_{ij} = \beta_{0j} + \beta_{1j}t_{ij} + \beta_2 t_{ij}^2 + e_{ij}$ was fitted. To avoid correlation between the *t* terms, orthogonal polynomials were used. While they have the advantage of being uncorrelated, their estimated coefficients are not directly interpretable, therefore γ_{10} and γ_{20} in Table i above cannot be compared with the coefficient in Appendix 3, Table 4.D.

Figure i: Scatterplot of observed PISA scores versus scores predicted by Model C, by domain.



Eventually, Model C was replaced by one that took into consideration the clustering of countries into OECD and partner economies (Model E, Table ii).

Table ii: A model using OECD as a Level-2 (time-invariant) predictor (Model E)

	Model E Reading	**Model E Maths**	**Model E Science**
γ_{00} (intercept)	414.73*** (9.36)	427.31*** (10.68)	437.80*** (8.60)
γ_{01} (*OECD*)	75.83*** (12.84)	68.10*** (14.79)	61.01*** (11.96)
γ_{10} (*t*)	6.23*** (1.16)	5.80*** (1.34)	4.20*** (1.23)
γ_{11} (*t*OECD*)	-5.18*** (1.46)	-6.10*** (1.74)	-2.77* (1.64)
σ^2_{u0}	2536 (1728–3722)	3517 (2457–5035)	2327 (1636–3308)
σ^2_{u1}	14.01 (5.61–34.97)	31.68 (17.33–57.92)	28.29 (16.22–49.33)
σ_{u01}	-127.84	-209.59	-76.64
σ^2_e	103.71 (81.45–132.06)	44.04 (32.47–59.74)	24.47 (16.91–35.39)

***p <0.01, **p <0.05, *p <0.1

The final step was to introduce a variable capturing access to pre-primary education. One possibility was to add this predictor *x* to the Level-2 unit *j*, but it would have implied that it is constant over time, whereas one of the main purposes of this paper is to see if changes in the indicators are associated to changes in PISA scores. On the other hand, since the *x* varies from year to year, it could be entered as a Level-1 variable in the equation, but this would have caused other issues:

> 'Though the *x* variable of interest varies both over time and across countries, these two dimensions of its variability are treated singly—making it impossible to know whether just one dimension is driving any covariation found with *y*, or even potentially whether the two associations have different signs. This approach effectively assumes that the cross-sectional and longitudinal relationships between *x* and *y* are the same: a single coefficient β linked to [x_j] captures both.' (Fairbrother, 2013, p. 10)

Following Fairbrother (2013) and others (Enders & Tofighi, 2007; Raudenbush & Bryk, 2002; Raudenbush, 1989), a solution was adopted by decomposing the indicators into two components—a time-invariant, group-centred mean at Level-2 ($\bar{x}_j$), and its Level-1, time-varying difference (x^*_{ij}), as follows:

$$x_{ij} = \bar{x}_j + x^*_{ij} \quad (7)$$

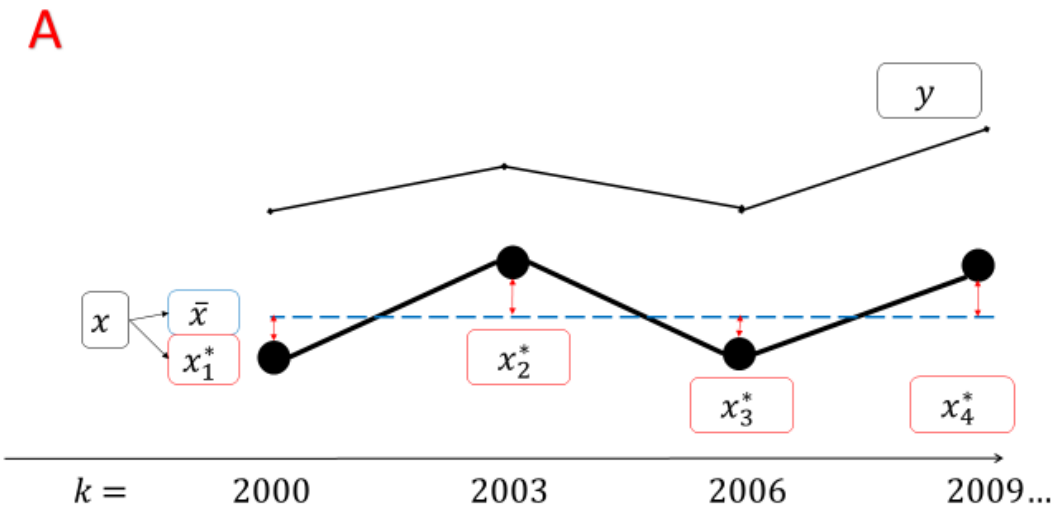
$$x^*_{ij} = x_{ij} - \bar{x}_j \quad (8)$$

Once all amendments were taken into account, Equation (4) became:

$$Scores_{ij} = \gamma_{00} + \gamma_{01}\bar{x}_j + \gamma_{02}OECD_j + \gamma_{10}t_{ij} + \gamma_{20}t_{ij}OECD_j + \gamma_{20}x^*_{ij} + (u_{0j} + + u_{1j}t_{ij} + e_{ij}) \quad (9)$$

Equation (9) represents Model F, and the decomposition of the predictor measure is summarised in Figure ii.

Figure ii: Decomposition of the predictor into two orthogonal components



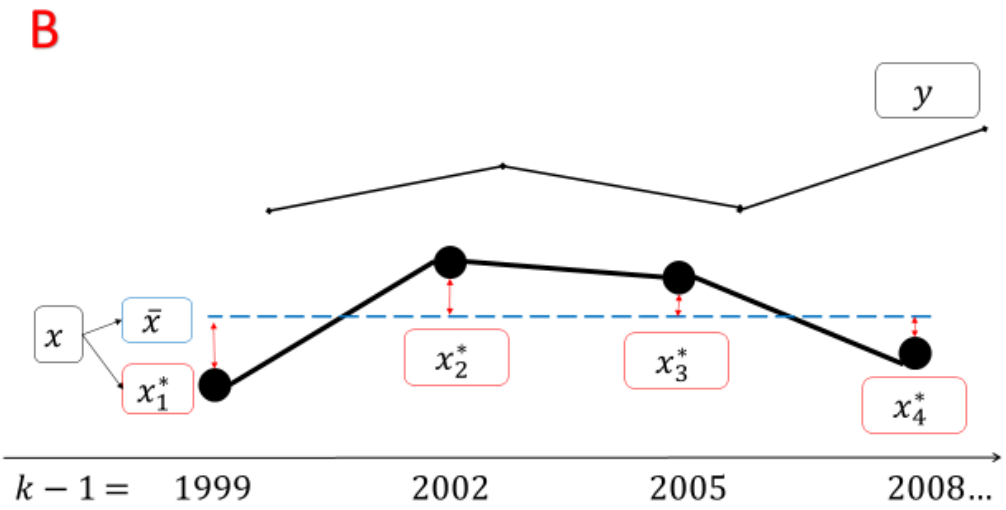
Since $\sum_{i=1}^{N}(x_i - \bar{x}) = 0$, $\bar{x}_j$ and x_{ij}^* are orthogonal and the effect of x_{ij}^* is independent from the magnitude of $\bar{x}_j$. In other words, this model allows us to explore changes in the predictor variable regardless of the average conditions in each country. Equation (9) relies on the assumption that such effect would be equal in each country. Normally, this assumption would be unwarranted, but the reader is reminded that in this research x_{ij}^* captures the effect of changes in the regressors *over and above* the effect of time within a relatively short time frame: therefore, its value is expected to be rather small and consistent across Level-2 units.

The use of a lagged predictor

Equation (9) provides a general model, but consider once again the relationship displayed in Figure ii. What the model is testing is whether and how changes in predictor variables are related to simultaneous changes in the outcome variable within the same time frame. Since some variables used in this research, such as enrolment in pre-primary education, must be considered at the time when it could have had an impact on the PISA cohort, the time frame of the predictor was shifted so that PISA scores were compared to past values.

A shift of one year (k) corresponds to one lag unit (λ), so that $k_x = k_y - \lambda$. For instance, to model the relationship between changes in the predictor one year before PISA, and PISA scores, the time frame of **x** is shifted by one year ($\lambda = 1$), so that $Scores_{k\ =\ \{2000,\ 2003,\ 2006,\ 2009,\ 2012\}}$ is regressed on $x_{k-1\ =\ \{1999,\ 2002,\ 2005,\ 2008,\ 2011\}}$ (Figure iii).

Figure iii: Shifting of the time frame of the predictor variable



The amount of shifting depends on the predictor (see the section on variables). Shifting the time frame also affects the group-centred mean, $\bar{X}$: this makes it possible to understand the evolution of Level-2 effects over time; the trade-off is that the size of the Level-1 effects cannot be compared over time, as it is relative to the group-centred mean.

Outcome variable

The outcome variable *Scores* is the country mean estimates in Reading, Mathematics, and Science published by OECD in its international reports (OECD, 2014g, and previous). These estimates are measured on a scale with a mean of 500 and a standard deviation (SD) of 100, and they are calculated from the students' plausible values[29] (OECD, 2014h). It is worth remembering that country estimates are associated to an error component that is generally small (between one and three PISA points) but which can reach six or seven points for some non-OECD countries, especially in the early cycles. This means that the 95% confidence interval of a country whose estimate was reported to be 500 could actually range from 486 to 514 points.

On a cross-sectional level, countries can be separated by hundreds of points. For instance, 245 points divided the top and bottom scorers in Mathematics 2012 (Shanghai, 613; Peru, 368); the average score in that year was 473, the median 483, the SD 56 points and the median absolute deviation about 35. From a longitudinal perspective, however, score changes from

[29] Plausible values are not scores, but a function representing 'a likely distribution of a student's proficiency' (von Davier, Gonzalez, & Mislevy, 2009, p. 11). Student-level analysis of PISA data generally requires the use of a subset of five such values that were obtained through random draws from each student's posterior distribution (OECD, 2009d) and can be found in the PISA databases.

one cycle to the next are much smaller. It can be shown that most country score changes are well below ±20 points (Lenkeit & Caro, 2014, calculated that 95% of them are between -13.8 and +17.8 points), and although larger improvements or drops have been recorded, their being reflective of actual student ability has been questioned on the grounds of the PISA equating methodology and the differential functioning of certain items in some countries (see for example Gebhardt & Adams, 2007; Monseur & Berezner, 2007; Wu, 2010).

Input variables

Time

Time enters this research in three forms: as a coded variable for each PISA administration (or measurement occasion, *t*); as calendar year (*k*); and as a measure of lag (λ). The measurement occasion *t* orders the time-points in which each assessment was held and, in order to measure growth, it is zeroed at the first comparable PISA in the trend. Since different domains have their first comparable test at different times, the same value of *t* can stand for different calendar years *k*: therefore, $t = 0$ represents the year 2000 ($k = 2000$) in case of Reading; but $t = 0$, $k = 2003$, for Mathematics; and $t = 0$, $k = 2006$, in the case of Science. As explained above, the lag λ is the number of years by which a certain time-point or time frame can be shifted back in time.

Economic productivity

Two variables were used to measure economic productivity: GDP per capita (variable: *GDP*) expressed in constant 2005 US Dollars, and income group. *Income* is a categorical variable that uses The World Bank division of economies into low, lower-middle, upper-middle, and high income according to gross national income (GNI) per capita. *Income* takes two values: *lower*, which stands for countries in the low, lower-middle and upper-middle income groups; and *high*. The *high* category corresponds to a GNI per capita above 9,265 USD for PISA 2000 and 12,615 USD during the 2012 assessment (The World Bank, 2014). *Income* was used to cluster countries and extract subsamples in some analyses. [Sources: UNESCO Institute for Statistics, The World Bank, and International Monetary Fund].

The choice regarding the amount of lag to apply was based mainly on practical constraints, as the literature on systemic change provides little insights on the time required for countries to transform their wealth and expenditures for education into effective policies. Time frames range from two years (Earl et al., 2002), to four (Fullan, 2007), five (Borman et al., 2003), eight (Fullan, 2000), ten (Harris, 2011) until up to a generation (Wedell, 2009). The stability of PISA scores suggests that, at the global level, change is small and consistent, with countries improving by about 1 point per year in Reading, or even less in the case of Mathematics and

Science. Clear trends emerge after three or four administrations (or 9 to 12 years). This could be evidence that reforms take a decade to make a sizeable impact or that PISA can capture long-lasting systematic effects but not rapid changes in education systems. In either case, it seemed best to focus on changes in *GDP* and *Income* that had taken place during the lifetime of each PISA cohort.

Measuring equity

Equity is measured in terms of the Strength and the Slope of the socio-economic gradient (OECD, 2013e, p. 35; Willms, 2006). The "socio-economic gradient" is the regression line (or curve) of student scores on their composite index of economic, social and cultural status (ESCS). It has different attributes (see OECD, 2013e, p. 35; Willms, 2006), but the most important are its Strength and its Slope (capital letters are used to differentiate them from general statistical terms).

The "Strength" of the socio-economic gradient is the amount of score variance captured by the regression line (in statistical terms, the R^2): the higher the "explained" variance, the stronger the ESCS-score relationship, and the less equitable a system. The "Slope", instead, refers to the extent to which differences in socio-economic background are associated to differences in outcomes, with steeper slopes being related to greater disparities. Theoretically, only one between the Strength and Slope would be necessary for analysing country equity, because if ESCS and PISA scores were converted to standardised scores (Z-scores), the correlation between them would be the Slope of the line and the R^2 the Strength. Nevertheless, it was decided to follow the OECD procedure and run analyses using both attributes as the outcome variable.

The PISA database does not have socio-economic gradients for all administrations, but up-to-date ESCS indexes were provided on the OECD website to enable comparisons over time. Therefore, using the methodology detailed in the *Data Analysis Manual* (OECD, 2009d), a new dataset with comparable values for the Strength (variable: *R-squared*) and the Slope (*Slope*) of the socio-economic gradient was created and it is available upon request. Since these variables are constructed from PISA data they are entered in the model without any lag (conceptually, a country has the same amount of time to alter both the educational and the equity outcomes of its students).

Measuring the expansion of pre-primary education

The variable *GER* is the gross enrolment ratio in pre-primary education. It is calculated as the ratio, expressed as a percentage, between the number of children enrolled in ECE and the number of all children in the official pre-primary age group.

ECE generally targets children 'from age 3 years to the start of primary education' (UNESCO-UIS, 2012, p. 26). By accessing UNESCO databases, it was established that the average primary school starting age was 6.3 years for all PISA 2000–2012 participants. Therefore, *GER* was measured when the population was between three and five years of age, which corresponded to time lags of 10, 11 and 12 from each PISA cycle. To obtain a unique value to correlate to each PISA score, it was decided to estimate *GER* when the average child would have been around four (λ = 11) by calculating the average *GER* of lags 10–12. The resulting time frame is summarised in Table iii.

Table iii: Time frame for gross enrolment ratio in pre-primary education (GER)

Lag	Student age (average)	GER	Outcome	Time points
10/12	4	1988/90–2000/02	R 2000–2012	5
10/12	4	1991/93–2000/02	R, M 2003–2012	4
10/12	4	1994/96–2000/02	R, M, S 2006–2012	3

Since some countries are flexible about enrolment in ECE of children of a different age from the reference group, reported GER values can sometimes be higher than 100%.

Accounting for OECD membership

OECD is a dummy variable that was entered both as a Level-2 unit (to represent its contribution to variations in the intercept), and as a Level-1 interaction term with time (to represent its effect on the slope, see Equation 9). It takes the value of 1 for OECD countries and 0 for partner countries. Three countries (Chile, Estonia and Israel) joined OECD only in 2010, but since the process to acquire membership must have begun some years earlier and they were all high-income economies before or by 2012 (The World Bank, 2014), they were considered OECD members for the purpose of this paper.

Appendix 6: Model diagnostics

All models rely on assumption concerning the fixed and random components. In this paper, a framework for model diagnostics was adapted from Pinheiro & Bates (2000). The first assumption concerns the Level-1 error components and is formally represented by Equation 5, whereas the second assumption concerns the Level-2 components and is represented by Equation 6.

1) The Level-1 components are:
    a. Independent from one another
    b. Normally distributed
    c. With mean 0 and variance σ^2
2) The Level-2 components are:
    a. Independent from one another
    b. Normally distributed
    c. With mean 0 and covariance matrix Ω.

The growth model in Appendix 3, Table 4.D, and represented by Equation 4 in Appendix 5, was analysed through this framework. While the following sections only apply to the Reading domain (because it has the greatest number of time-points and shows the higher variability), the same tests can be carried out (and similar considerations be made) for Mathematics and Science.

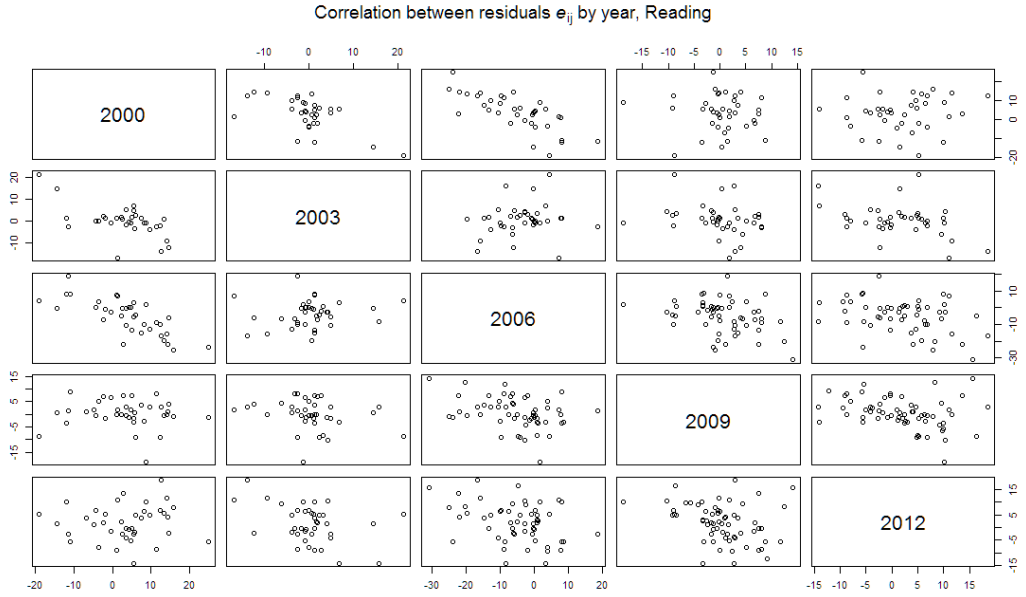
Level-1 components

Assumption of independence

Given that at each time point the observations came from the same 'individuals' (countries), the assumption of independence between residuals may be unwarranted, and their autocorrelation could be instead a better choice. Autocorrelation in Level-1 residuals happens when the errors in two consecutive observations for the same individual j, e_{1j} and e_{2j} are not independent from one another. Therefore, their covariance is not 0.

To investigate into this possibility, residuals from the same country from different years were plotted against one another (Figure iv).

Figure iv: Matrix plot of within-country, between-years correlations. Each dot represents a country.



Most correlations appear to be rather close to 0, yet the correlation between two consecutive years (e.g., 2006–2009 or 2009–2012) or that between 2000 and 2006 seem stronger and negative. This impression was put to the test by analysing the outcome of the empirical autocorrelation function [(ACF), Pinheiro & Bates, 2000], which takes this form:

$$\hat{\rho}(\lambda) = \frac{\sum_{j=1}^{M}\sum_{i=1}^{n_j-\lambda} r_{ij}r_{(i+l)j}/N(\lambda)}{\sum_{j=1}^{M}\sum_{i=1}^{n_j} r_{ij}^2/N(0)} \quad (10)$$

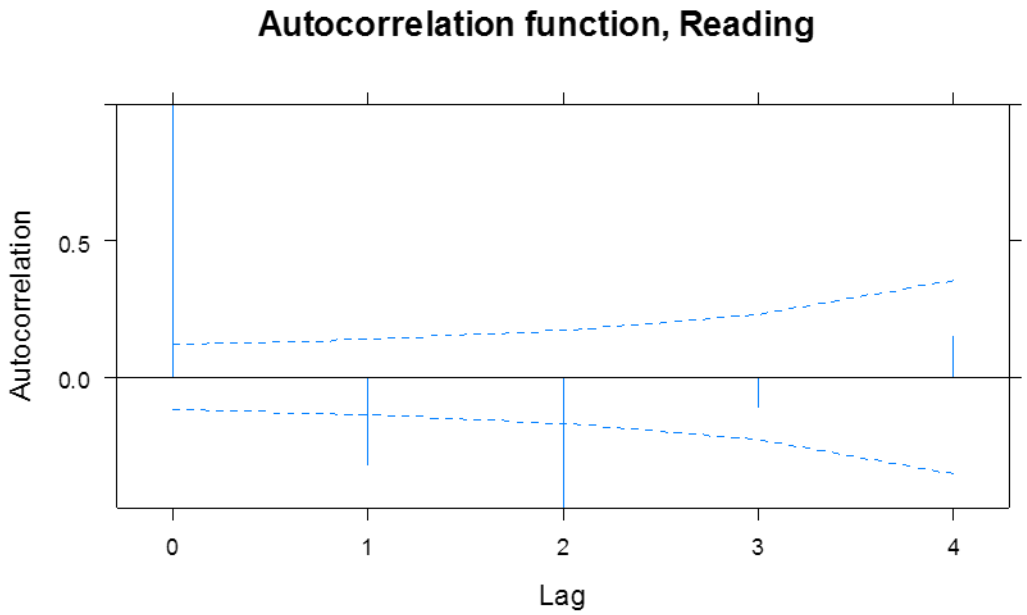
Where:

- $\hat{\rho}(\lambda)$ is the estimated correlation parameter at lag λ
- M is the number of Level-2 groups, indexed by j[30]
- n_j is the number of observations in group j
- r_{ij} are the standardised Level-1 residuals returned by a fitted multilevel model
- $N(\lambda)$ is the number of residual pairs used at the given lag

Following this equation, $\hat{\rho}$ was calculated and plotted (Figure v), and the initial impression of negatively correlated residuals at lag 1 and 2 was confirmed.

[30] Note that Pinheiro & Bates (2000) follow an opposite notation, indexing Level-2 variables with *i* and Level-1 variables with *j*.

Figure v: Autocorrelation function for Level-1 residuals. The dashed lines represent the 0.05 significance level.



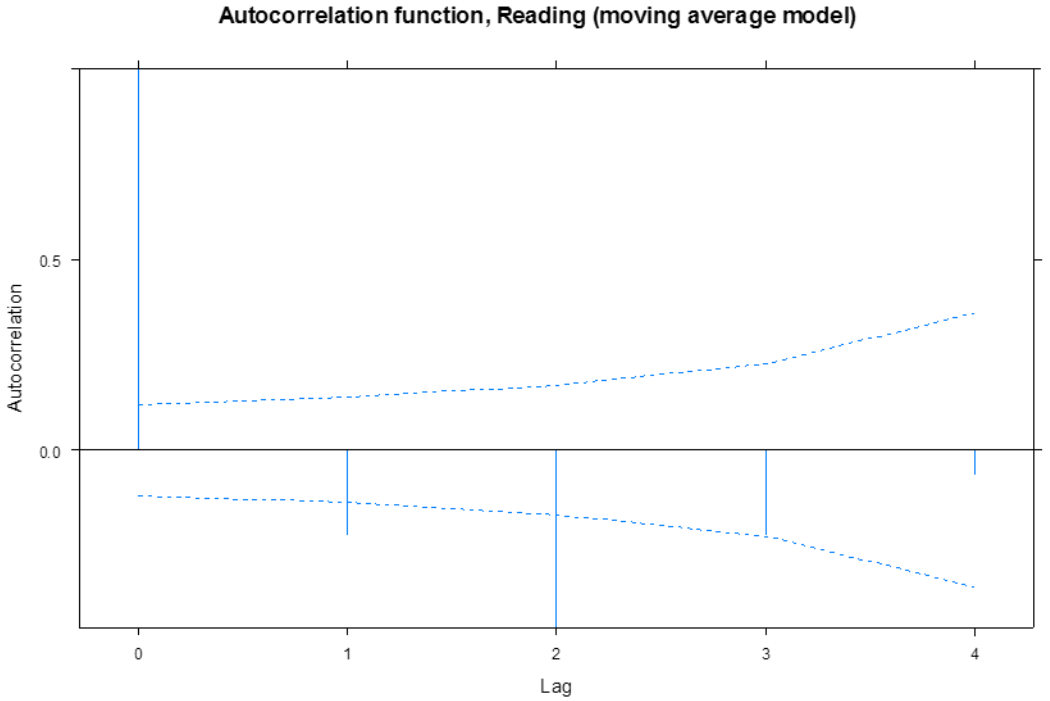
Since there is evidence of autocorrelation for the first two lag, we analysed models of the form ARMA(*p*,*q*) (Box, Jenkins, & Reinsel, 2011), with *p* and/or *q* = {0,1,2}, as possible explanatory candidates. The analyses showed that both the ARMA(1,0) and the ARMA(0,1) models represented a slight improvement over the baseline model, with this latter providing the best alternative (Table iv).

Table iv: Comparison between the baseline and the autoregressive moving average model.

Model	DF	AIC	BIC	logLik	Lik.Ratio	p-value
Baseline growth	6	2310.932	2332.319	-1149.466		
Autoregressive [AR(1)]	7	2310.143	2335.095	-1148.072	2.78856	0.0949
Baseline growth	6	2310.932	2332.319	-1149.466		
Moving average [MA(1)]	7	2309.550	2334.501	-1147.775	3.382225	0.0659

Despite it being significant at the 0.1 level, the MA(1) model does not explain all of the residual correlation as one would expect if it were a plausible fit, but tends instead to reduce the correlation at lag 1 and to increase it at lag 3 (Figure vi).

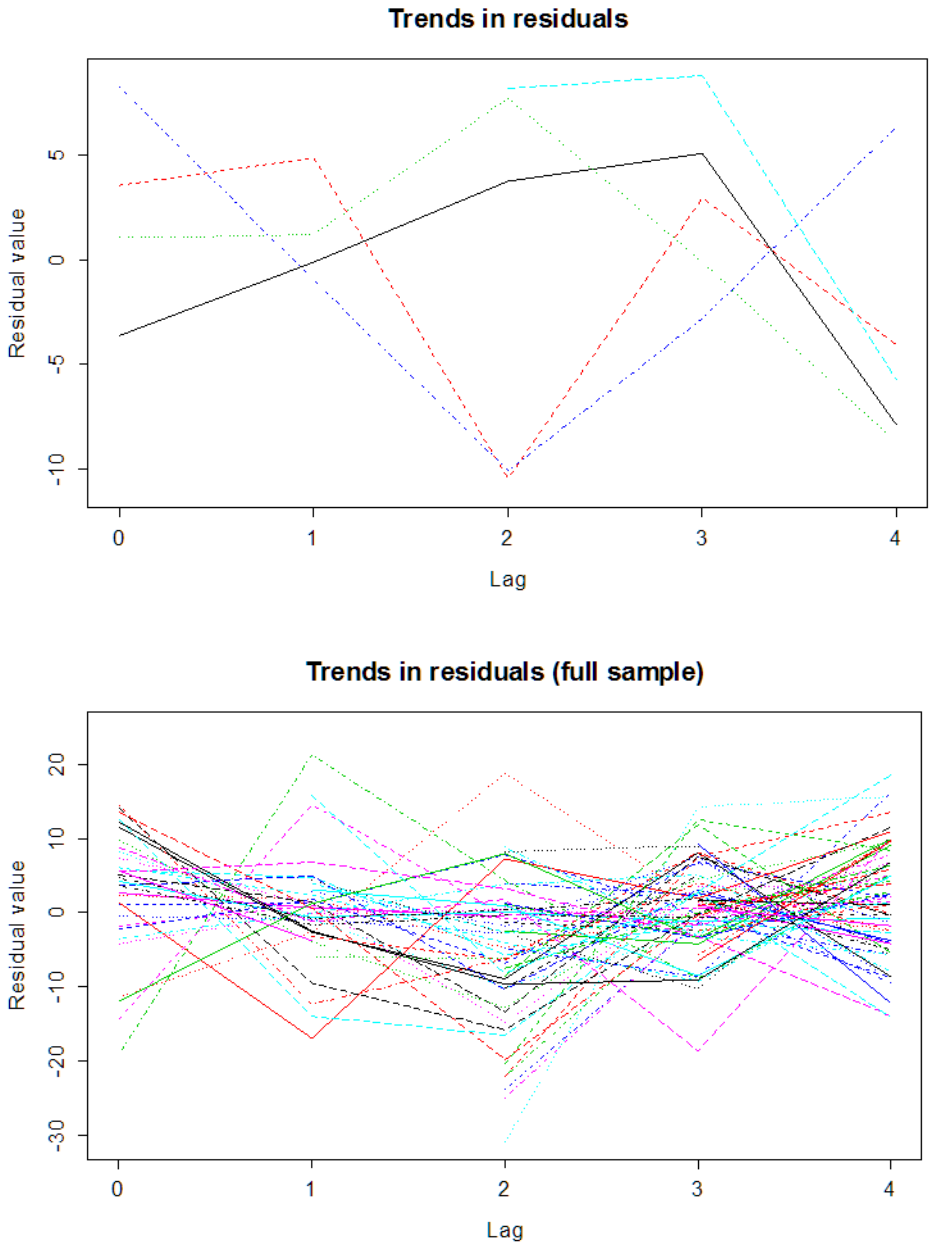
Figure vi: Autocorrelation function for Level-1 residuals using an ARMA(0,1) model. The dashed lines represent the 0.05 significance level.



To understand why this may be the case, recall that the empirical correlation function in Equation A aggregates within-country results at the country level by computing the sum of the Level-1 products across all groups. This means that while each individual country may not present any evidence of autocorrelation, the trends in Level-1 residuals of all countries taken together might; in other words, PISA countries may show what may be called 'emergent' or 'apparent' autocorrelation.

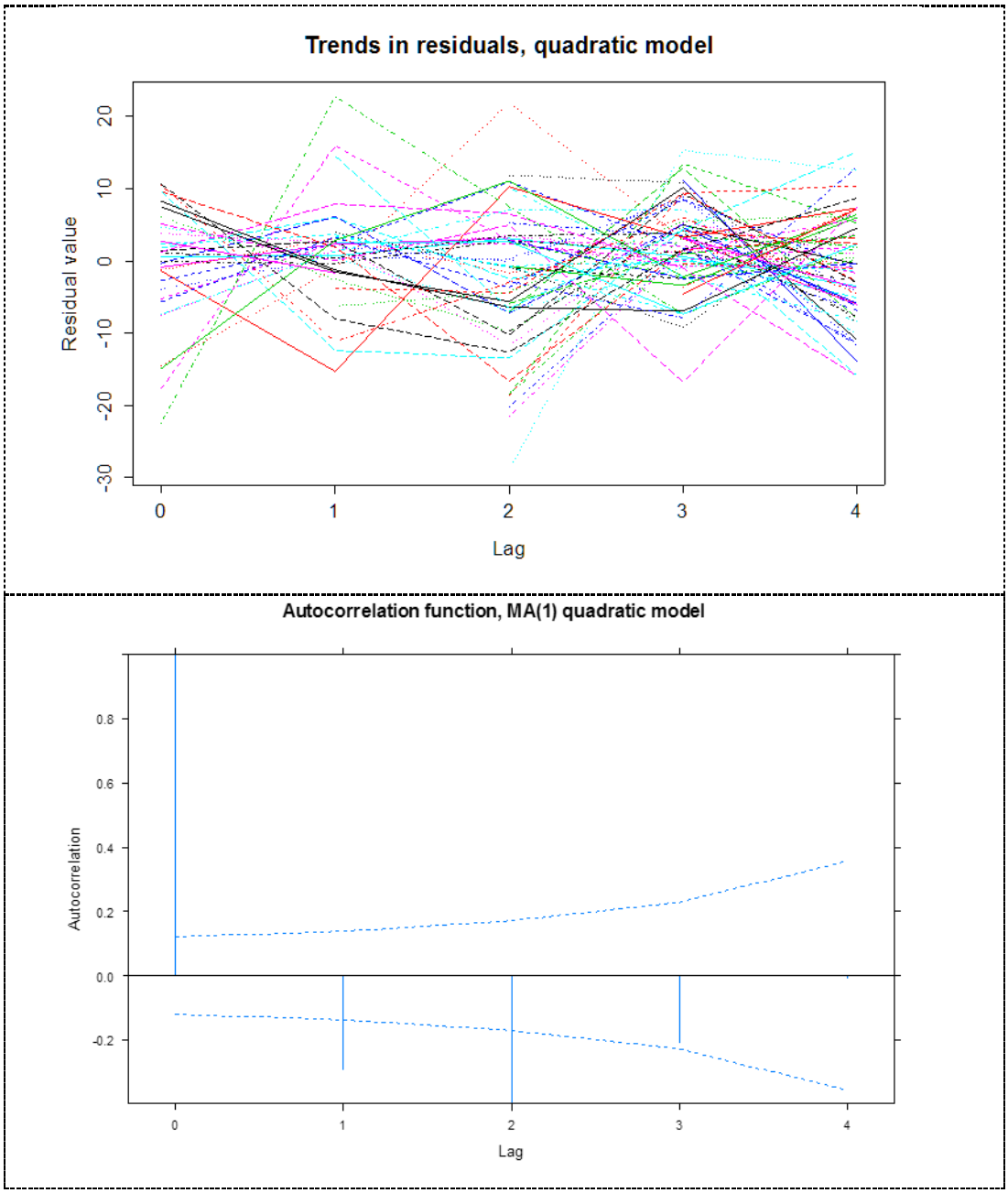
Support for this hypothesis can be found by looking at the trends in residuals. Figure vii compares the trends in residuals of a random sub-sample of five countries from the dataset ($\mathbf{D}_3$, at the top) to those of the full sample ($\mathbf{D}_0$, at the bottom). Trends from $\mathbf{D}_3$ look indeed random: for instance, $r_{2,\text{RED}}$ and $r_{2,\text{VIOLET}}$ are lower than $r_{0,\text{RED}}$ and $r_{0,\text{VIOLET}}$, whereas the opposite is true for $r_{2,\text{GREEN}}$ and $r_{2,\text{BLACK}}$; all residuals show a decrease from lag 3 to lag 4, but $r_{4,\text{VIOLET}}$ does not; etc. In the global trends from $\mathbf{D}_0$ represented at the bottom, instead, there are hints of an emergent pattern, whereby *most* residuals seem to be decreasing between $\lambda = 0$ and $\lambda = 2$.

Figure vii: Comparison between the residual trends for a random sub-sample of five countries in the dataset (top) and the trends of the full sample (bottom).



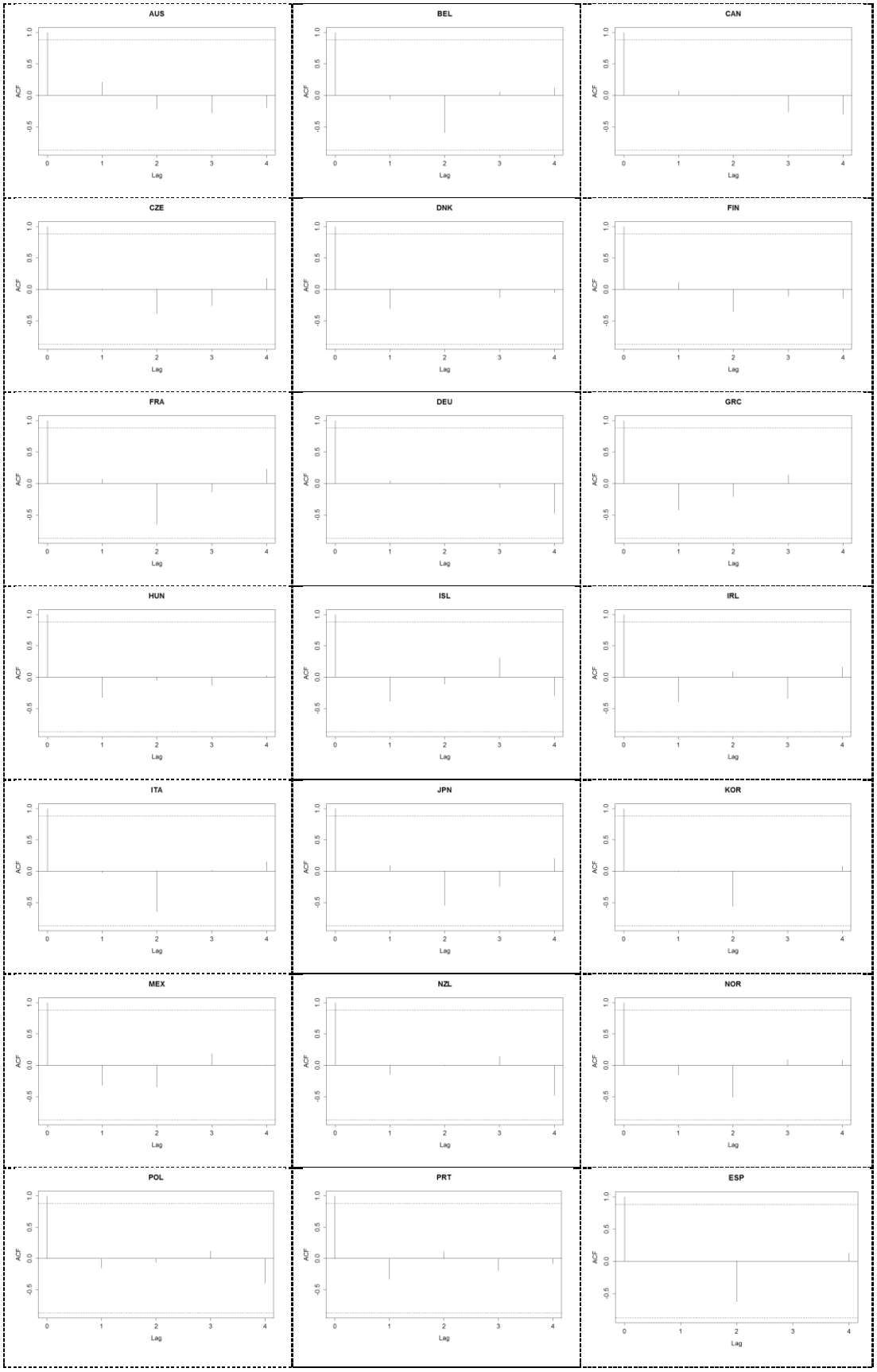
Notice that trends in residuals from $\mathbf{D}_0$ seem to follow a curvilinear trajectory, and indeed a quadratic growth model for Reading can be fitted (Figure viii). While this slightly 'straightens' the residual pattern, however, it does not eliminate the autocorrelation, which suggests that the causes for values of $\rho \neq 0$ must lie elsewhere—for instance in the regression towards the mean effect noted in the main text.

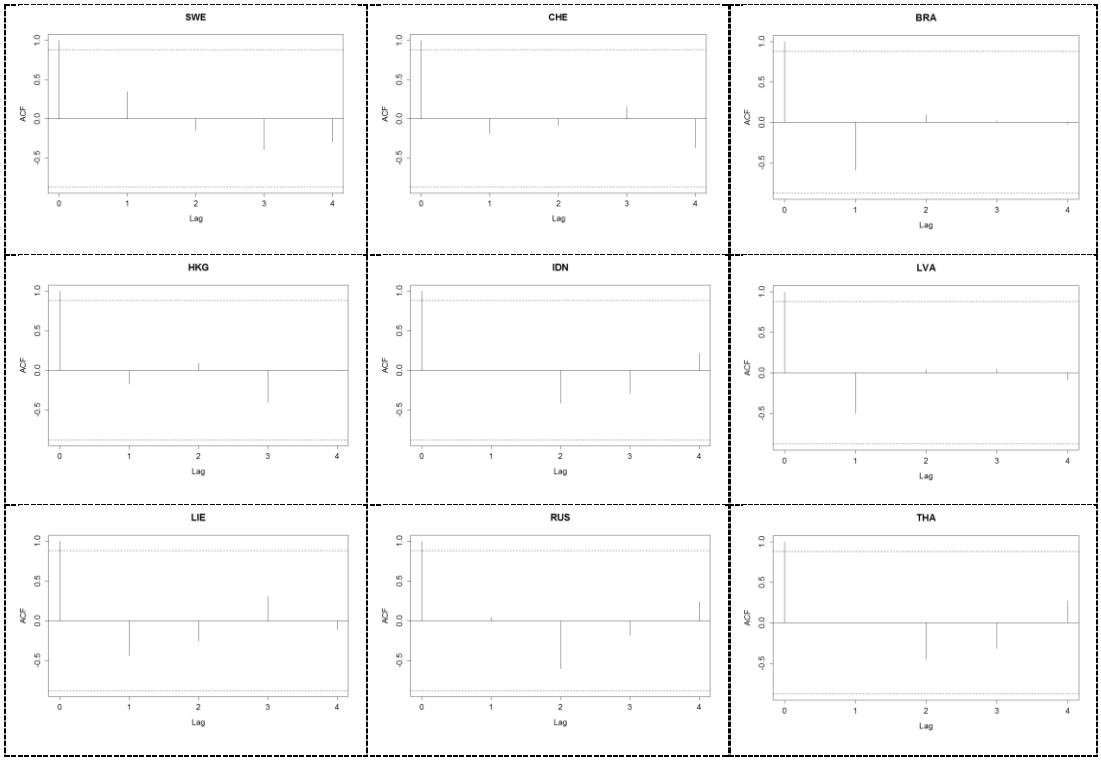
Figure viii: Trends in residual and ACF graph for a quadratic growth model.



Overall, these findings give backing to the idea that while some global trends in PISA results can be identified, differences among countries will be more evident than their similarities until more data are produced, making in-depth analyses worthwhile only at the national level. Figure ix completes this picture by representing the ACF for each member of $\mathbf{D}_1$ (that is, the dataset containing only countries with valid observations at all time-points): in no instance can the autocorrelation be said to be statistically different from 0, suggesting that the base growth model using measurement occasion as a predictor captures well both within- and between-country score variation.

Figure ix: Autocorrelation function for countries with complete observations. The dashed lines represent the 0.05 significance level.

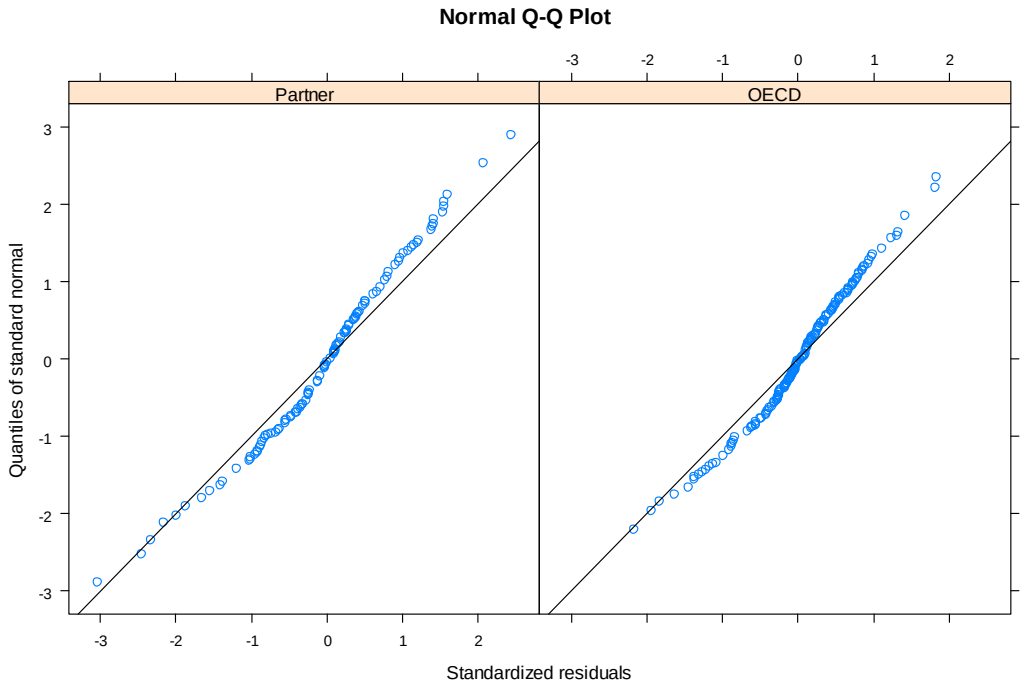




Assumption of normality

An examination of a normal Q-Q plot for OECD and partner countries (Figure x) shows that the distribution of the Level-1 residuals does not deviate too much from normality, although there tend to be more observations close to the mean than there are in a normal distribution.

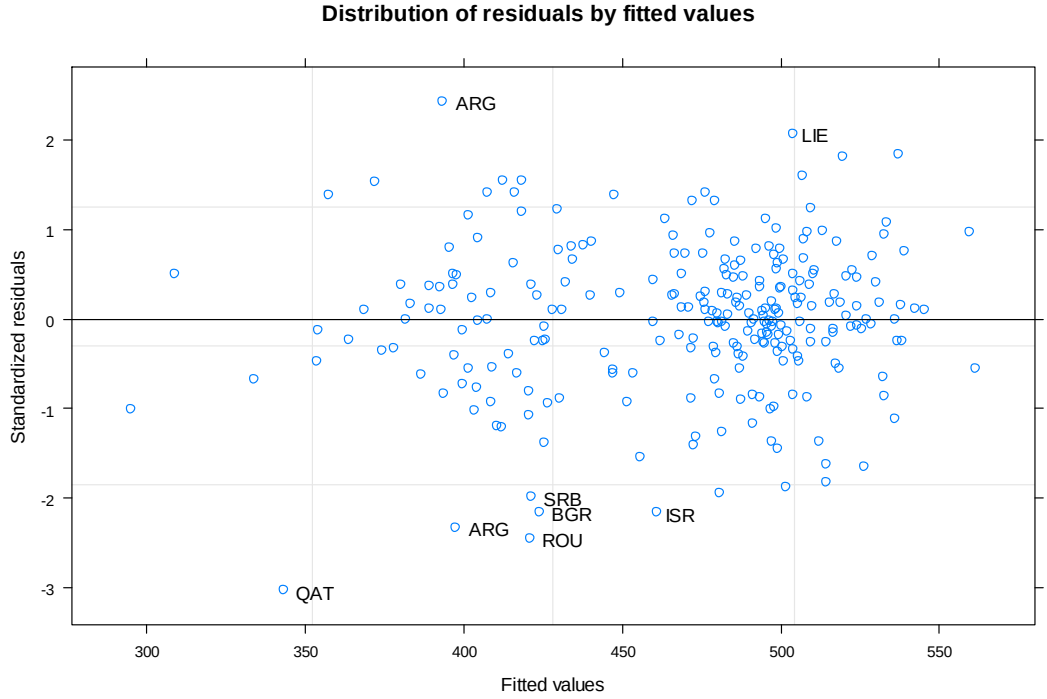
Figure x: Normal Q-Q plot for Level-1 residuals.



Assumptions about the mean and the variance

In single-level regression, a routine diagnostic test to check whether residuals have a mean 0 and equal variance σ^2 is a scatterplot of standardised residual against fitted values—and the one in Figure xi shows that, despite some outlying observations, the assumption of homoscedasticity seems to be well-founded.

Figure xi: Distribution of residuals by fitted values.



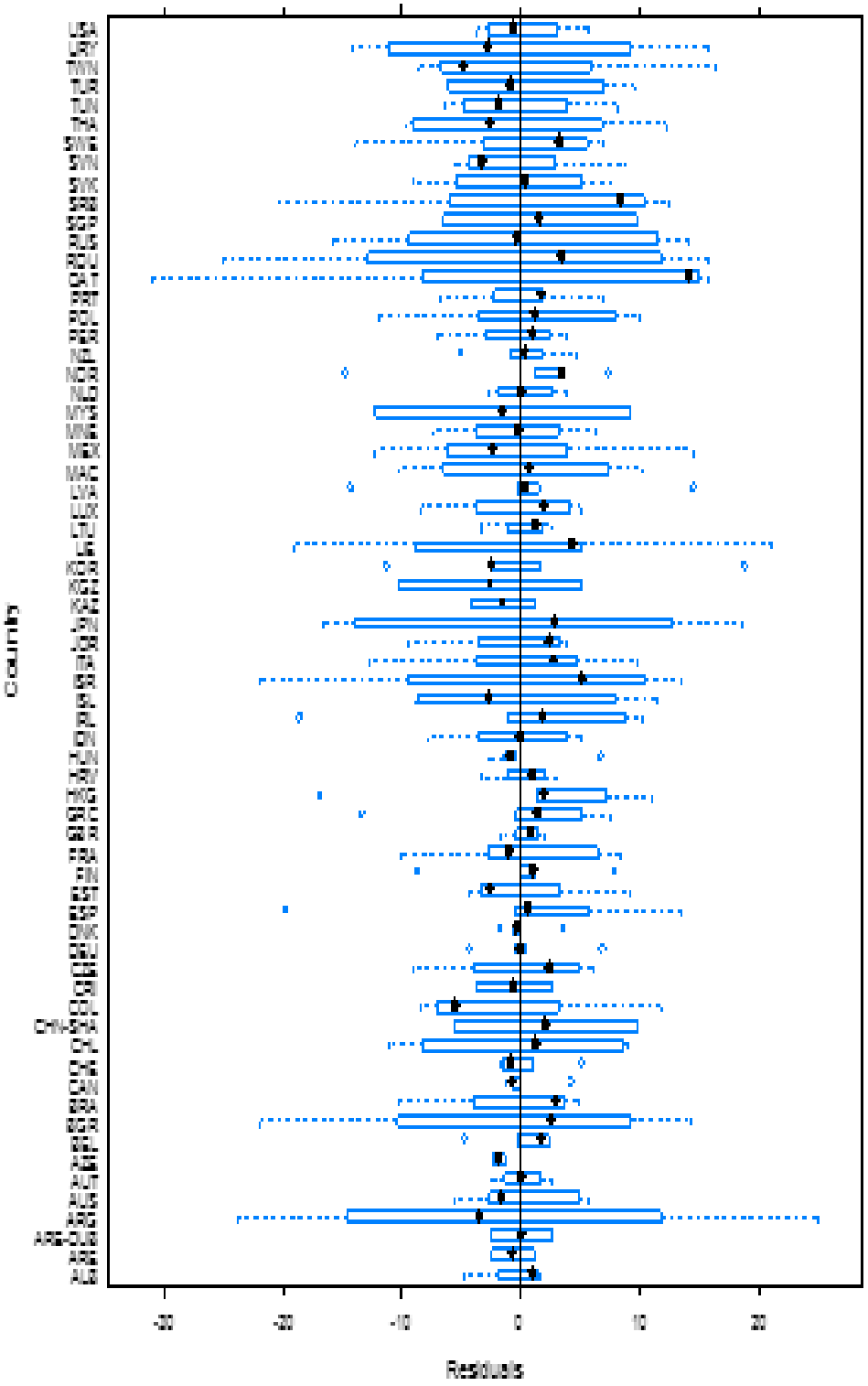
This procedure, however, is less optimal when used with nested data, because it can mask important differences between groups. For instance, note how Argentina appears twice in Figure xi, and at opposite sides of the graph, which suggests a wider spread that most other countries. To assess the within-country variance, another plot taking grouping into account was produced (Figure xii). While the expectation of the mean can be said to be approximately 0—with the exception of outliers such as Qatar or Serbia—there is clearly greater between-country variation in within-country variance, with trends such as that of the United Kingdom being almost perfectly captured by the growth model and others, such as that of Argentina, which are instead more unpredictable.

Ideally, one would want to model each country's variance individually, but this would result in an over-parameterised model and would also defeat the purpose of multilevel modelling. Another approach, instead, is to divide countries into a manageable number of categories and weight the variance according to them. A compromise was found, in this case, by splitting the variance component into two, one for OECD countries and one for partner countries, whereby $\sigma^2_{OECD} \cong 0.648\sigma^2_{Partner}$.

Figure xiii shows how individual residuals are affected by the weighting: the change in conditions entails that different countries are identified as outliers.

Given that the non-weighted and weighted models are nested, their explanatory power can be compared with a likelihood ratio test. Table v shows that the weighted model provides a better fit for the data, and highlights how variances are affected by the transformation. While both the Reading and Mathematics models can be adapted in this fashion, Science data are mostly homoscedastic and do not benefit from a different variance function.

Figure xii: Within-country variance and mean.



The boxplots represent the estimated spread of residuals, whereas the black dot the estimated mean.

Figure xiii: Change in the Level-1 variance after applying a weight based on OECD membership.

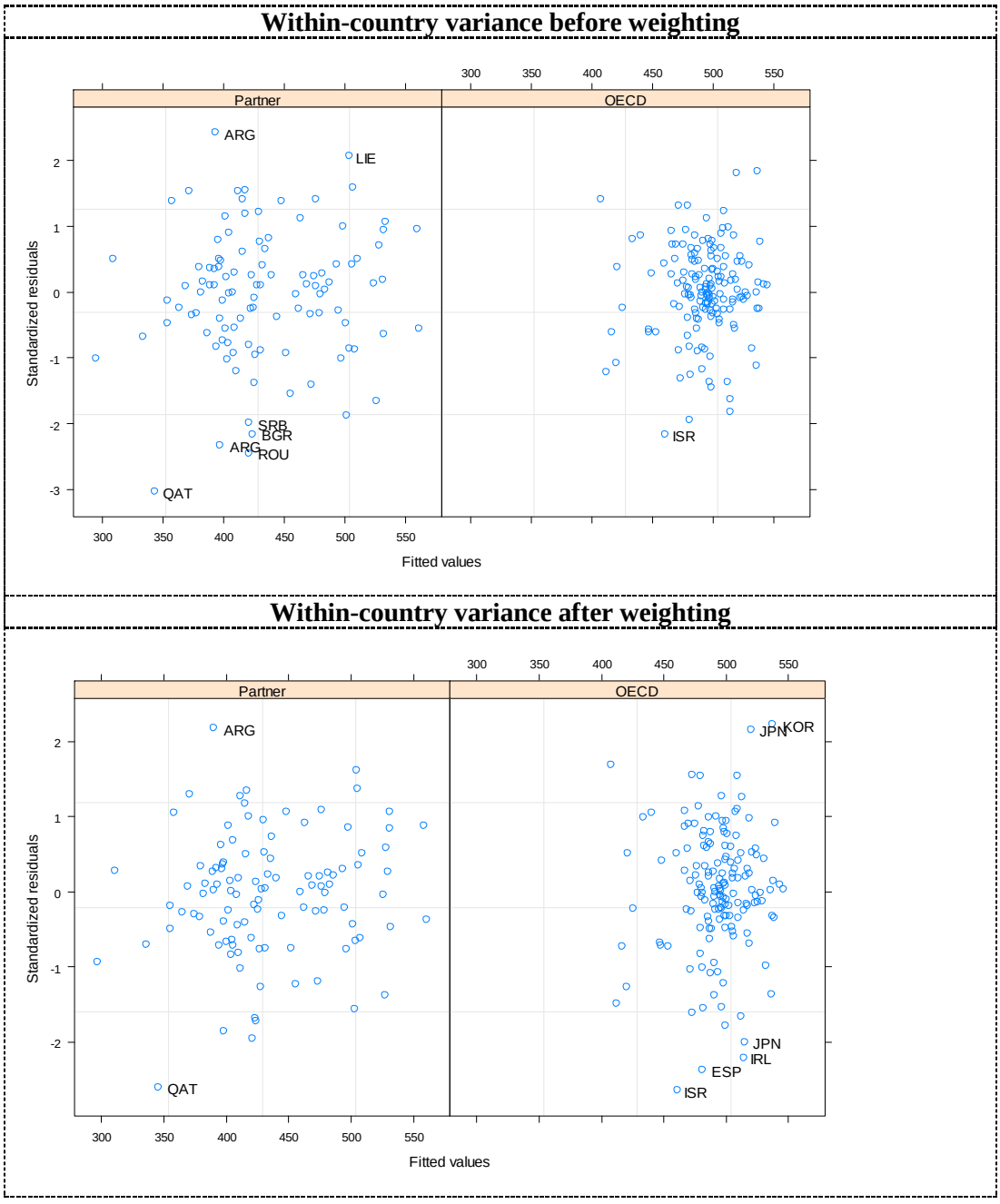


Table v: Comparison between the non-weighted and weighted model, Reading.

	Before weighting	After weighting
DF	6	7
AIC	2310.932	2298.031
BIC	2332.319	2322.983
Log-likelihood	-1149.466	-1142.015
Likelihood ratio		14.90116
p-value		<0.01
σ^2_{u0}	4012* (2785–5780)	3993* (2762—5774)
σ^2_{u1}	20.45* (9.51–43.97)	19.53* (9.57—39.85)
σ^2_{e}	104.31* (81.95–132.78)	166.40* (116.71—237.24)

Level-2 components

Assumption of normality

Q-Q plots can be analysed to check for significant departures from normality of the level-2 residuals. In this case, separate plots can be made for the intercept and the time residuals, as well for OECD countries and partner countries (Figure xiv and Figure xv).

Figure xiv: Q-Q plot of Level-2 residuals.

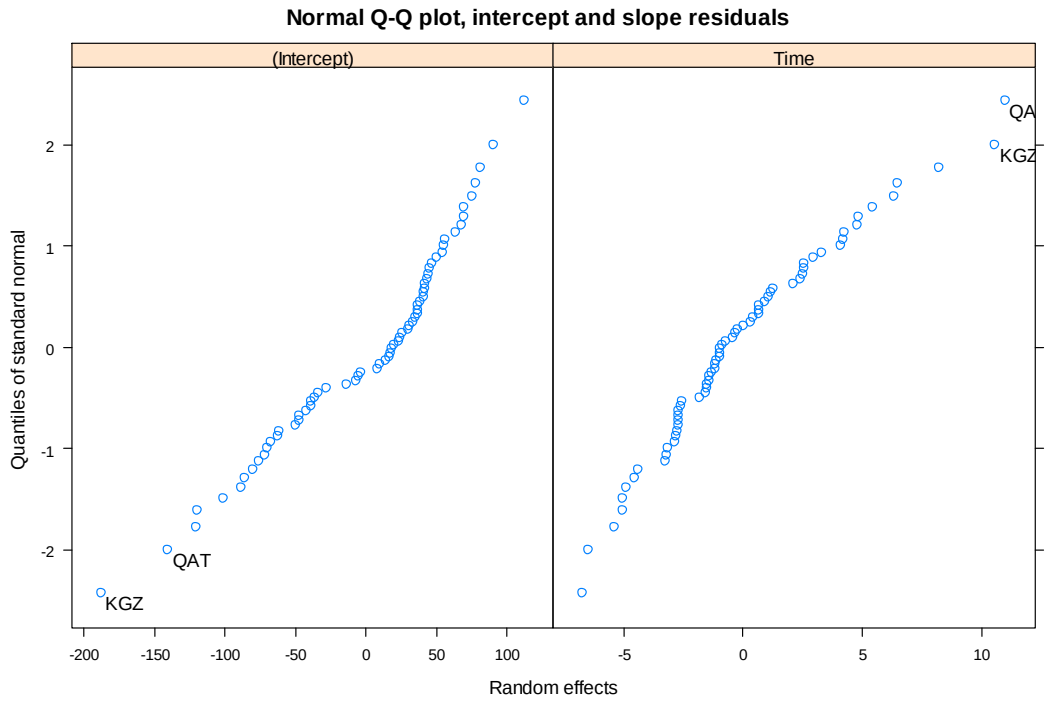
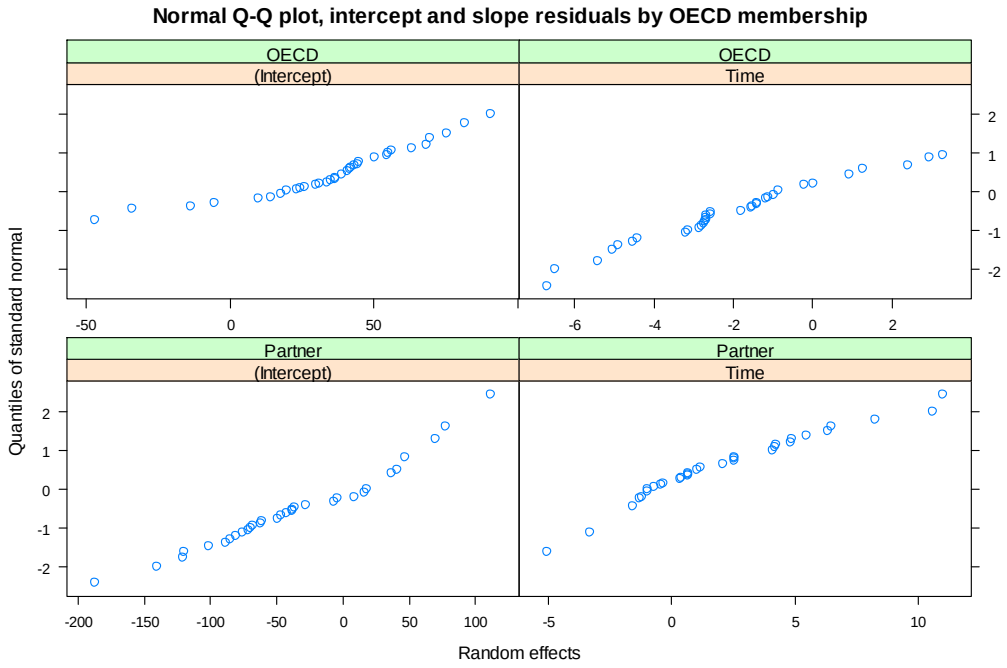


Figure xv: Q-Q plot of Level-2 residuals by OECD membership.

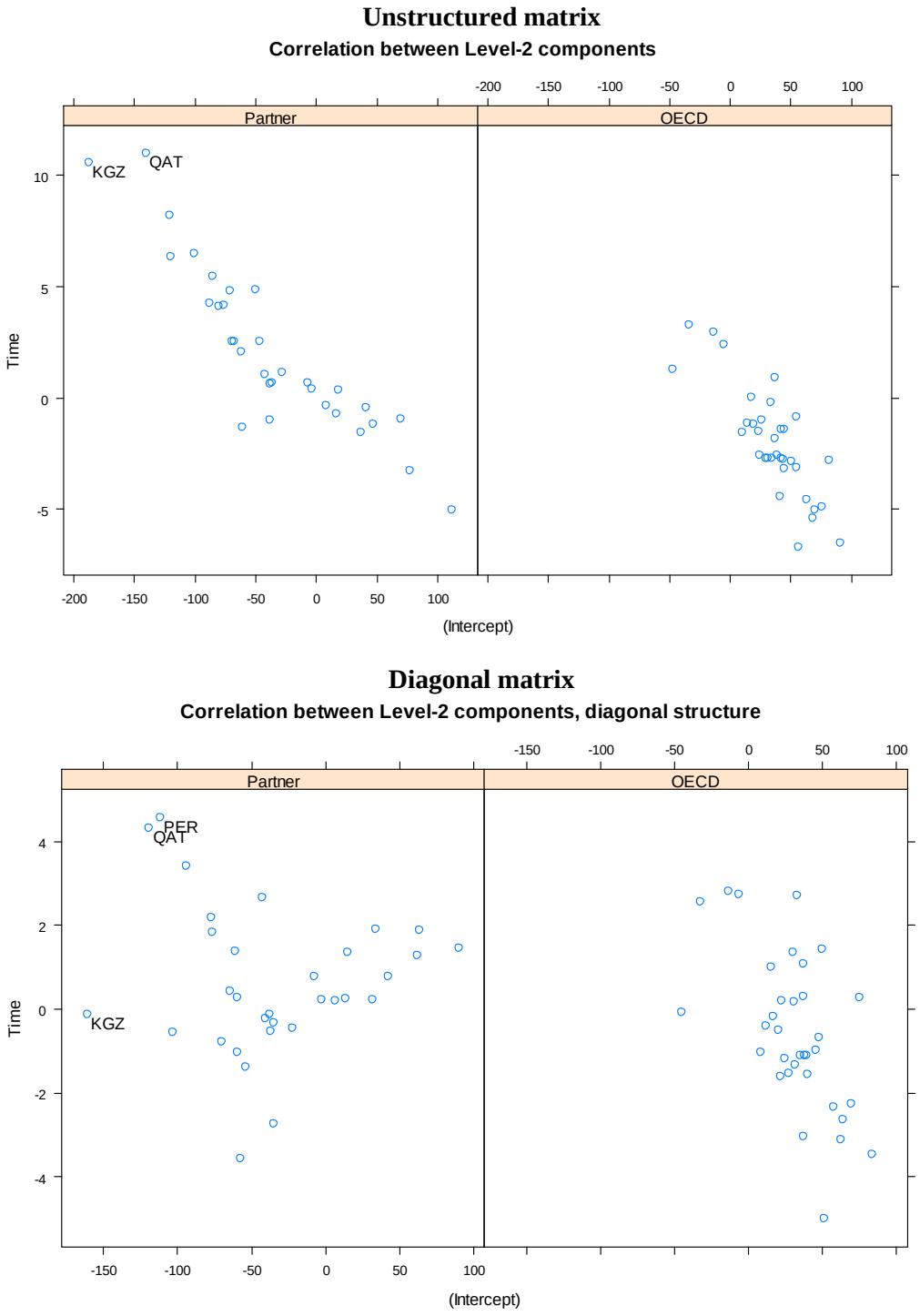


Once again, there are sine departures from normality but they are not as extreme to warrant a complete redefinition of the model.

Assumption of homogeneity

The top panel of Figure xvi plots the strong negative correlation between estimated intercept and slope. This is a result of the general positive-definite variance matrix of the model, which imposes no structure on the variance components. By looking at the actual trends over time, however, this modelling could exaggerate actual covariances. To test this possibility, the baseline growth model was compared with one using a diagonal matrix that forced $\sigma_{u01} = 0$. Doing so produces different variance estimates, and the new estimated covariance between intercept and time residuals is displayed in the bottom panel of Figure xvi.

Figure xvi: Correlation between intercept and slope by OECD membership.



Since the two models are nested, their informative power can be compared through a likelihood ratio test (Table vi). In this case, unstructured model proves to be a better fit for the data by all information criteria.

Table vi: Analysis of variance of two models, one with an unstructured covariance matrix, one assuming 0 covariance

Model	DF	AIC	BIC	logLik	Lik.Ratio	p-value
Diagonal $(\sigma_{u01} = 0)$	5	2332.937	2350.760	-1161.469		
Baseline growth $(\sigma_{u01} \neq 0)$	6	2310.932	2332.319	-1149.466	24.00516	<0.01

It could be argued that OECD and partner countries should be modelled independently, as their scores seem to vary and be affected by time in substantially different ways. While this is probably true, once this splitting is done, other dissimilarities within these two groups could be found, and one would have no reason not to continue in splitting groups into smaller entities until a single country is tracked in its progression since PISA 2000. In other words, the purpose of multilevel modelling is to provide a compromise between an analytic and a synthetic, or holistic, inquiry. Besides, while OECD members have effectively something in common (OECD membership), non-OECD membership cannot be said to be a *trait d'union* of partner countries, which share indeed very little: it would be therefore simplistic to treat them as a homogenous group.

References

Aberbach, J. D., & Rockman, B. A. (2002). Conducting and Coding Elite Interviews. *PS: Political Science and Politics*, *35*(4), 673–676. http://doi.org/10.2307/1554807

Adams, R. J. (2003). Response to "Cautions on OECD"s Recent Educational Survey (PISA)'. *Oxford Review of Education*.

Adams, R. J., & Wu, M. L. (Eds.). (2002). *PISA 2000 Technical Report*. Paris: OECD Publishing.

Adey, P., Hewitt, G., Hewitt, J., & Landau, N. (2004). *The Professional Development of Teachers: Practice and Theory*. Dordrecht: Kluwer Academic Publishers.

Akour, M., Sabah, S., & Hammouri, H. (2015). Net and Global Differential Item Functioning in PISA Polytomously Scored Science Items: Application of the Differential Step Functioning Framework. *Journal of Psychoeducational Assessment*, *33*(2), 166–176. http://doi.org/10.1177/0734282914541337

Alasuutari, P., & Rasimus, A. (2009). Use of the OECD in justifying policy reforms: the case of Finland. *Journal of Power*, *2*(1), 89–109. http://doi.org/10.1080/17540290902760899

Allen, J. P., Hafen, C. A., Gregory, A. C., Mikami, A. Y., & Pianta, R. (2015). Enhancing Secondary School Instruction and Student Achievement: Replication and Extension of the My Teaching Partner – Secondary Intervention. *Journal of Research on Educational Effectiveness*. http://doi.org/10.1080/19345747.2015.1017680

Almond, D., & Currie, J. (2010). *Human Capital Development Before Age Five*. Retrieved from http://www.nber.org/papers/w15827

Aloisi, C. (2014). OECD Policy Advice Implementation Questionnaire. Retrieved April 16, 2016, from http://community.dur.ac.uk/cesare.aloisi/OECD_PAIQ.html

Ammermueller, A. (2013). Institutional Features of Schooling Systems and Educational Inequality: Cross-Country Evidence From PIRLS and PISA. *German Economic Review*, *14*(2), 190–213. http://doi.org/10.1111/j.1468-0475.2012.00565.x

Antoniou, P., & Kyriakides, L. (2011). The impact of a dynamic approach to professional development on teacher instruction and student learning: results from an experimental study. *School Effectiveness and School Improvement*, *22*(3), 291–311. http://doi.org/10.1080/09243453.2011.577078

Apple, M. W. (1988). Social Crisis and Curriculum Accords. *Educational Theory*, *38*(2), 191–201. http://doi.org/10.1111/j.1741-5446.1988.00191.x

Araújo, L., & Dinis da Costa, P. (2012). *Differential Item Functioning (DIF): What Functions Differently for Immigrant Students in PISA 2009 Reading Items?* Luxembourg: Publications Office of the European Union.

Archer, P., & Shortt, F. (2003). *Review of the home/school/community liaison scheme*. Dublin: Educational Research Centre.

Archer, P., & Weir, S. (2004). *Addressing Disadvantage: A review of the international literature and of strategy in Ireland*. Dublin: Educational Disadvantage Committee.

Areepattamannil, S. (2012). Effects of Inquiry-Based Science Instruction on Science Achievement and Interest in Science: Evidence from Qatar. *The Journal of Educational Research*, *105*(2), 134–146. http://doi.org/10.1080/00220671.2010.533717

Areepattamannil, S., Melkonian, M., & Khine, M. S. (2015). International note: Exploring differences in native and immigrant adolescents' mathematics achievement and dispositions towards mathematics in Qatar. *Journal of Adolescence*, *40*, 11–13. http://doi.org/10.1016/j.adolescence.2014.12.010

Armingeon, K. (2004). OECD and national welfare state development. In K. Armingeon & M. Beyeler (Eds.), *The OECD and European Welfare States* (pp. 226–241). Cheltenham: Edward Elgar.

Arts, B., Liefferink, D., Kamstra, J., & Ooijevaar, J. (2008). The Gap Approach: What Affects the Direction of Environmental Policy Convergence? In K. Holzinger, C. Knill,

& B. Arts (Eds.), *Environmental Policy Convergence in Europe: The Impact of International Institutions and Trade* (pp. 196–226). New York, NY: Cambridge University Press.
ASPE. (2013). *Décrochage et abandon scolaire précoce: Mise en perspective européenne de la situation en Fédération Wallonie – Bruxelles*. Bruxelles: Ministère de la Communauté française.
Assessment Reform Group. (1999). *Assessment for learning: beyond the black box*. Cambridge: School of Education, University of Cambridge.
Austin, J. L. (1962). *How to do things with words*. London: Oxford University Press.
Baird, J.-A., Isaacs, T., Johnson, S., Stobart, G., Yu, G., Sprague, T., & Daugherty, R. A. (2011). *Policy effects of PISA*. Oxford. Retrieved from http://research-information.bristol.ac.uk/files/14590358/Baird_et_al._2011.pdf
Baker, R., & Jones, A. (2005). How can international studies such as the international mathematics and science study and the programme for international student assessment be used to inform practice, policy and future research in science education in New Zealand? *International Journal of Science Education*. http://doi.org/10.1080/0950069042000276695
Ball, S. J. (1993). What Is Policy? Texts, Trajectories and Toolboxes. *Discourse: Studies in the Cultural Politics of Education*, *13*(2), 10–17. http://doi.org/10.1080/0159630930130203
Ball, S. J. (2015). Education, governance and the tyranny of numbers. *Journal of Education Policy*, *30*(3), 299–301. http://doi.org/10.1080/02680939.2015.1013271
Ball, S. J., Maguire, M., & Braun, A. (2012). *How Schools do Policy Policy Enactments in Secondary Schools*. New York, NY: Routledge. http://doi.org/10.4324/9780203153185
Barber, M., & Mourshed, M. (2007). *How the world's best-performing schools systems come out on top*. London: McKinsey & Co.
Barnett, M. N., & Finnemore, M. (1999). The Politics, Power, and Pathologies of International Organizations. *International Organization*, *53*(4), 699–732. http://doi.org/10.2307/2601307
Barnett, M. N., & Finnemore, M. (2004). *Rules for the World: International Organizations in Global Politics*. Ithaca, NY: Cornell University Press.
Barnett, W. S. (2011). Effectiveness of early educational intervention. *Science*, *333*, 975–978. http://doi.org/10.1126/science.1204534
Barrera-Osorio, F., Garcia-Moreno, V., Patrinos, H. A., & Porta, E. (2011). *Using the Oaxaca-Blinder Decomposition Technique to Analyze Learning Outcomes Changes over Time: An Application to Indonesia's Results in PISA Mathematics* (No. 5584). *Policy Research Working Paper*. The World Bank. Retrieved from http://elibrary.worldbank.org/content/workingpaper/10.1596/1813-9450-5584
Bates, D. M. (2009). *Assessing the precision of estimates of variance components* (Presentation at the LMU, Munich, 16th July). Retrieved from http://lme4.r-forge.r-project.org/slides/2009-07-16-Munich/
Baye, A., Chenu, F., Crahay, M., Lafontaine, D., & Monseur, C. (2014). *Le redoublement en Fédération Wallonie-Bruxelles: Rapport d'expertise commandité par le groupe parlementaire du Parti Socialiste*. Liège: Service d'Analyse des Systèmes et des Pratiques d'Enseignement, Université de Liège. Retrieved from http://orbi.ulg.ac.be/bitstream/2268/165801/1/Le redoublement en FWB.pdf
Baye, A., Demonty, I., Fagnant, A., Matoul, A., & Monseur, C. (2004a). *PISA 2003* (Presentation). Retrieved from http://www.enseignement.be/download.php?do_id=3053&do_check=
Baye, A., Demonty, I., Fagnant, A., Matoul, A., & Monseur, C. (2004b). *PISA 2003: quels défis pour notre système éducatif* (Summary report). Retrieved from http://www.enseignement.be/download.php?do_id=2761&do_check=
Baye, A., Demonty, I., Fagnant, A., Matoul, A., & Monseur, C. (2005). PISA 2003: au-delà des moyennes, des constats qui forcent à l'action. *Tables Rondes*, *1*(1), 2–5.

http://doi.org/10.1787/9789264010543-en
Baye, A., Demonty, I., Lafontaine, D., Matoul, A., & Monseur, C. (2010). La lecture à 15 ans. Premiers résultats de PISA 2009. *Les Cahiers Des Sciences de l'Éducation*, *31*.
Baye, A., Fagnant, A., Hindryckx, G., Lafontaine, D., Matoul, A., & Quittre, V. (2009). Les compétences des jeunes de 15 ans en Communauté française en sciences, en mathématiques et en lecture. Résultats de l'enquête PISA 2006. *Cahiers Des Sciences de l'Education*, *29-30*.
Baye, A., & Monseur, C. (2016). Gender differences in variability and extreme scores in an international context. *Large-Scale Assessments in Education*, *4*(1). http://doi.org/10.1186/s40536-015-0015-x
Beaton, A. E., Martin, M. O., Mullis, I. V. S., Gonzalez, E. J., Smith, T. A., & Kelly, D. L. (1996). *Science Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study*. Chestnut Hill, MA: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.
Beaton, A. E., Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., Kelly, D. L., & Smith, T. A. (1996). *Mathematics Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.
Beck, V., Fuller, A., & Unwin, L. (2006). Safety in stereotypes? The impact of gender and "race" on young people's perceptions of their post-compulsory education and labour market opportunities. *British Educational Research Journal*, *32*(5), 667–686. http://doi.org/10.1080/01411920600895718
Becker, G. S. (1964). *Human capital: A theoretical and empirical analysis with special reference to education*. New York, NY: National Bureau of Economic Research.
Becker, G. S. (1993). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education* (3rd ed.). Chicago, IL: University of Chicago Press.
Becker, M., Neumann, M., Tetzner, J., Böse, S., Knoppick, H., Maaz, K., … Lehmann, R. (2014). Is early ability grouping good for high-achieving students' psychosocial development? Effects of the transition into academically selective schools. *Journal of Educational Psychology*, *106*(2), 555–568. http://doi.org/10.1037/a0035425
Beckers, J. (2007). Le métier change, la formation aussi: Comment contribuer par la formation des agrégés aux réformes du système éducatif. *Puzzle: Bulletin Du Centre Interfacultaire de Formation Des Enseignants*, (22), 4–10.
Beckers, J. (2011). Contribuer en Communauté française de Belgique à une approche par compétences qui réduirait les inégalités d'apprentissage. *Puzzle: Bulletin Du Centre Interfacultaire de Formation Des Enseignants*, (30), 9–15. Retrieved from http://www.ulg.ac.be/cms/c_3186161/fr/publications
Beckers, J., Florkin, É., & Leroy, J. (2008). Table ronde: Les valeurs sous-jacentes au Décret-missions et au contrat pour l'école. *Puzzle: Bulletin Du Centre Interfacultaire de Formation Des Enseignants*, (25), 11–12. Retrieved from http://www.ulg.ac.be/cms/c_3186161/fr/publications
Beckers, J., & Voos, M.-C. (2003). *Reforms of the teaching profession: a historical survey (1975-2002) - French community of Belgium* (Country report No. EC-31-12-850-EN-N). European Commission. Eurydice European Unit. Retrieved from http://bookshop.europa.eu/en/key-topics-in-education-in-europe-pbEC3112850/?CatalogCategoryID=AGwKABstsBgAAAEjpJEY4e5L
Belfield, C. R., & Levin, H. M. (2002). The Effects of Competition between Schools on Educational Outcomes: A Review for the United States. *Review of Educational Research*, *72*(2), 279–341. http://doi.org/10.2307/3516034
Belfield, C. R., & Levin, H. M. (2005). Vouchers and Public Policy: When Ideology Trumps Evidence. *American Journal of Education*, *111*(4), 548–567. http://doi.org/10.1086/431183
Bengtsson, J. (2008). *OECD's Centre for Educational Research and Innovation — 1968 to 2008* (Paper presented at the OECD/CERI International Conference "Learning in the

21st Century: Research, Innovation and Policy"). Paris: OECD Publishing.
Benton, T. (2014). Using meta-regression to explore moderating effects in surveys of international achievement. *Practical Assessment, Research & Evaluation*, *19*(3). Retrieved from http://pareonline.net/pdf/v19n3.pdf
BERA. (2011). *Ethical Guidelines for Educational Research* (2nd rev.).
Berliner, D. C. (2011). The context for interpreting PISA results in the USA. In M. A. Pereyra, H.-G. Kotthoff, & R. Cowen (Eds.), *PISA under examination: Changing knowledge, changing tests, and changing schools* (pp. 77–96). Rotterdam: SensePublishers.
Bernstein, B. (1975). *Class and Pedagogies: Visible and Invisible*. Washington, DC: OECD Publications Center.
Best, M., Knight, P., Lietz, P., Lockwood, C., Nugroho, D., & Tobin, M. (2013). *The impact of national and international assessment programmes on education policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries* (Final Report). London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.
Betts, J. R. (2011). The economics of tracking in education. In E. A. Hanushek, S. Machin, & L. Wößmann (Eds.), *Handbook of the Economics of Education* (Vol. 3, pp. 341–381). Amsterdam: North-Holland.
Beyeler, M. (2004). Introduction: A Comparative study of the OECD and European Welfare States. In K. Armingeon & M. Beyeler (Eds.), *The OECD and European Welfare States* (pp. 1–12). Cheltenham: Edward Elgar. http://doi.org/10.4337/9781843769705.00007
Bhreathnach, N. (1993). Adjournment Matters: Adult Literacy. *Seanad Éireann Debate*. 138(8), 25 November 1993.
Bhreathnach, N. (1994). Ceisteanna-Questions: Oral Answers - White Paper on Education. *Dáil Éireann Debate*. 445(8), 13 October 1994.
Bieber, T., & Martens, K. (2011). The OECD PISA Study as a Soft Power in Education? Lessons from Switzerland and the US. *European Journal of Education*, *46*(1), 101–116. http://doi.org/10.1111/j.1465-3435.2010.01462.x
Biebricher, T. (2015). Neoliberalism and Democracy. *Constellations*, *22*(2), 255–266. http://doi.org/10.1111/1467-8675.12157
Bimrose, J. (2009). *Careers guidance, identity and development*. Retrieved from http://www.beyondcurrenthorizons.org.uk/careers-guidance-identity-and-development/
Binder, M. (2009). Why Are Some Low-Income Countries Better at Providing Secondary Education? *Comparative Education Review*, *53*(4), 513–534. http://doi.org/10.1086/603641
Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 7–74. http://doi.org/10.1080/0969595980050102
Black, S. E., & Devereux, P. J. (2010). *Recent Developments in Intergenerational Mobility*. Retrieved from http://www.nber.org/papers/w15889
Blank, R. K., & de las Alas, N. (2009). *Effects of teacher professional development on gains in student achievement*. Washington, DC: Council of Chief State School Officers. Retrieved from www.ccsso.org
Blaug, M. (1976). Reviewed Work: "Education, Inequality and Life Chances." *The Economic Journal*, *86*(343), 664–665. Retrieved from http://www.jstor.org/stable/2230836
Bloem, S. (2013). *PISA in Low and Middle Income Countries* (No. 93). *OECD Education Working Papers*. Paris: OECD Publishing.
Blondin, C., & Giot, B. (2011). *Cadres d'évaluation en vue d'améliorer les résultats scolaires: Rapport de base pour la Communauté française de Belgique*. Bruxelles: Ministère de la Communauté française. Retrieved from http://www.dri.cfwb.be/index.php?id=1811

Boarini, R., & Lüdemann, E. (2009). The role of teacher compensation and selected accountability policies for learning outcomes: An empirical analysis for OECD countries. *OECD Journal: Economic Studies*, *2009*, 211–230. http://doi.org/10.1787/eco_studies-v2009-art10-en

Boddy, J., Neumann, T., Jennings, S., Morrow, V., Alderson, P., Rees, R., & Gibson, W. (2010). Limits of confidentiality: professional and elite interviews. Retrieved from http://www.ethicsguidebook.ac.uk/Limits-of-confidentiality-elite-interviews-232

Bol, T., & van de Werfhorst, H. G. (2013). Educational Systems and the Trade-Off between Labor Market Allocation and Equality of Educational Opportunity. *Comparative Education Review*, *57*(2), 285–308. http://doi.org/10.1086/669122

Bol, T., Witschge, J., Van de Werfhorst, H. G., & Dronkers, J. (2014). Curricular Tracking and Central Examinations: Counterbalancing the Impact of Social Background on Student Achievement in 36 Countries. *Social Forces*, *92*(4), 1–28. http://doi.org/10.1093/sf/sou003

Bolívar, A. (2011). The dissatisfaction of the losers: PISA public discourse in Ibero-American countries. In M. A. Pereyra, H.-G. Kotthoff, & R. Cowen (Eds.), *PISA under examination: Changing knowledge, changing tests, and changing schools* (pp. 61–74). Rotterdam: SensePublishers.

Bonal, X., & Tarabini, A. (2013). The Role of PISA in Shaping Hegemonic Educational Discourses, Policies and Practices: the case of Spain. *Research in Comparative and International Education*, *8*(3), 335–341. http://doi.org/10.2304/rcie.2013.8.3.335

Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.

Borman, G. D., Hewes, G. M., Overman, L. T., & Brown, S. (2003). Comprehensive School Reform and Achievement: A Meta-Analysis. *Review of Educational Research*, *73*(2), 125–230. http://doi.org/10.3102/00346543073002125

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, *111*(4), 1061–71. http://doi.org/10.1037/0033-295X.111.4.1061

Bowditch, C. (1993). Getting Rid of Troublemakers: High School Disciplinary Procedures and the Production of Dropouts. *Social Problems*, *40*(4), 493–509. http://doi.org/10.1525/sp.1993.40.4.03x0094p

Bowen, D. H., & Trivitt, J. R. (2014). Stigma Without Sanctions: The (Lack of) Impact of Private School Vouchers on Student Achievement. *Education Policy Analysis Archives*, *21*(12), 1–25. http://doi.org/10.14507/epaa.v22n87.2014

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2011). *Time series analysis: forecasting and control* (4th ed.). Hoboken, NJ: John Wiley & Sons, Inc.

Braga, M., Checchi, D., & Meschi, E. (2013). Educational policies in a long-run perspective. *Economic Policy*, *28*, 45–100. http://doi.org/10.1111/1468-0327.12002

Brauckmann, S., & Schwarz, A. (2014). Autonomous leadership and a centralised school system. *International Journal of Educational Management*, *28*(7), 823–841. http://doi.org/10.1108/IJEM-08-2013-0124

Breacháin, A. Ó., & O'Toole, L. (2013). Pedagogy or politics?: cyclical trends in literacy and numeracy in Ireland and beyond. *Irish Educational Studies*, *32*(4), 401–419. http://doi.org/10.1080/03323315.2013.851441

Breakspear, S. (2012). *The Policy Impact of PISA: An Exploration of the Normative Effects of International Benchmarking in School System Performance* (No. 71). *OECD Education Working Papers*. Paris: OECD Publishing.

Brist, L. E., & Caplan, A. J. (1999). More Evidence on the Role of Secondary Education in the Development of Lower-Income Countries: Wishful Thinking or Useful Knowledge? *Economic Development and Cultural Change*, *48*(1), 155–175. http://doi.org/10.1086/452450

Brown, G., Micklewright, J., Schnepf, S. V., & Waldmann, R. (2007). International surveys of educational achievement: how robust are the findings? *Journal of the Royal Statistical Society. Series A: Statistics in Society*, *170*(3), 623–646. Retrieved from

http://onlinelibrary.wiley.com/doi/10.1111/j.1467-985X.2006.00439.x/full
Brown, S., Ortiz-Nuñez, A., & Taylor, K. (2011). What will I be when I grow up? An analysis of childhood expectations and career outcomes. *Economics of Education Review*, *30*(3), 493–506. http://doi.org/10.1016/j.econedurev.2010.12.003
Brunello, G., & Checchi, D. (2007). Does school tracking affect equality of opportunity? New international evidence. *Economic Policy*, *22*(52), 782–861. http://doi.org/10.1111/j.1468-0327.2007.00189.x
Brunner, M., Artelt, C., Krauss, S., & Baumert, J. (2007). Coaching for the PISA test. *Learning and Instruction*, *17*(2), 111–122. http://doi.org/10.1016/j.learninstruc.2007.01.002
Bryman, A. (2006). Integrating quantitative and qualitative research: how is it done? *Qualitative Research*, *6*(1), 97–113. http://doi.org/10.1177/1468794106058877
Bryman, A. (2012). *Social Research Methods* (4th ed.). New York, NY: Oxford University Press.
Buchmann, C., DiPrete, T. A., & McDaniel, A. (2007). *Gender inequalities in education*. Retrieved from http://www.annualreviews.org/doi/abs/10.1146/annurev.soc.34.040507.134719
Byrne, D., & Smyth, E. (2010). *Behind the Scenes?: A Study of Parental Involvement in Post-primary Education*. Dublin: The Liffey Press.
Camilli, G., Vargas, S., Sharon, R., & Barnett, W. S. (2010). Meta-Analysis of the Effects of Early Education Interventions on Cognitive and Social Development. *Teachers College Record*, *112*(3), 579–620.
Campbell, T. (2015). Stereotyped at Seven? Biases in Teacher Judgement of Pupils' Ability and Attainment. *Journal of Social Policy*, *44*(3), 517–547. http://doi.org/10.1017/S0047279415000227
Cantillon, E. (2013). *Mixité sociale: le rôle des procédures d'inscription scolaire* (Working Paper). Retrieved from http://www.ecares.org/ecare/personal/cantillon/web/index.html
Carlson, D., Cowen, J. M., & Fleming, D. J. (2013). Life After Vouchers: What Happens to Students Who Leave Private Schools for the Traditional Public Sector? *Educational Evaluation and Policy Analysis*, *35*(2), 179–199. http://doi.org/10.3102/0162373712461852
Carroll, P. G. H., & Kellow, A. J. (2011). *The OECD: a study of organisational adaptation*. Cheltenham: Edward Elgar.
Cartwright, F. (2011). *PISA in Ireland, 2000-2009: Factors affecting inferences about changes in student proficiency over time*. Educational Research Centre. Retrieved from http://www.erc.ie/?p=229
Carvalho, L. M., & Costa, E. (2014). Seeing education with one's own eyes and through PISA lenses: considerations of the reception of PISA in European countries. *Discourse: Studies in the Cultural Politics of Education*, 1–9. http://doi.org/10.1080/01596306.2013.871449
Cattonar, B. (2010). Références transnationales et conception nationale des politiques: Le cas de la réforme de la formation des enseignants en Communauté française de Belgique. *Recherche et Formation*, *64*, 17–30.
Cattonar, B., Mangez, E., Delvaux, B., Mangez, C., & Maroy, C. (2009). *Réception, usage et circulation au niveau national d'un instrument supranational de régulation basé sur la connaissance: les enquêtes PISA - Le cas de la Communauté française de Belgique* (No. 12). Retrieved from www.knowandpol.eu
CEF. (1992a). *Avis relatif à la «Radioscopie de l'enseignement» et au «Rapport de l'O.C.D.E»* (Avis No. 4). Bruxelles: Conseil de l'Education et de la Formation.
CEF. (1992b). *Les objectifs généraux du système d'enseignement et de formation* (Avis No. 2). Bruxelles: Conseil de l'Education et de la Formation.
CEF. (1994a). *La discrimination positive, moyen de favoriser la réussite scolaire des enfants issus de milieux défavorisés* (Avis No. 20). Bruxelles: Conseil de l'Education et de la Formation.

CEF. (1994b). *Les cycles et l'évaluation formative comme moyens pédagogiques au service des objectifs de l'enseignement* (Avis No. 19). Bruxelles: Conseil de l'Education et de la Formation.
CEF. (1998a). *Avis n°54 du Conseil de l'Education et de la Formation relatif à l'évaluation des résultats des élèves en sciences (IEA-1995)* (Avis No. 54). Bruxelles: Conseil de l'Education et de la Formation.
CEF. (1998b). *Avis relatif à l'évaluation des résultats des élèves en mathématiques. (IEA - 1995).* (Avis No. 56). Bruxelles: Conseil de l'Education et de la Formation.
CEF. (1998c). *La promotion de la réussite des enfants issus de milieux défavorisés* (Avis No. 55). Bruxelles: Conseil de l'Education et de la Formation.
CEF. (2002). *Évaluation de la maîtrise des compétences des élèves: À propos du Programme International de l'OCDE pour le Suivi des acquis des élèves de 15 ans (PISA 2000)* (Avis No. 77). Bruxelles: Conseil de l'Education et de la Formation.
CEF. (2015a). *Etat des lieux du « Contrat pour l'Ecole »*. Bruxelles: Conseil de l'Education et de la Formation. Retrieved from http://www.cef.cfwb.be
CEF. (2015b). *Etat des lieux du « Contrat pour l'Ecole »: Tableau synthétique et chronologique* (No. http://www.cef.cfwb.be). Conseil de l'Education et de la Formation.
Centre de Recherche et d'Information Socio-Politique. (2015). Accord de la Saint Quentin. *Vocabulaire Politique*. Retrieved from http://www.vocabulairepolitique.be/accord-de-la-saint-quentin/
CERI. (1993). *Education at a Glance: OECD indicators* (2nd ed.). Paris: OECD Publishing.
CERI. (2008). *Learning in the 21st Century: Research, Innovation and Policy - Directions from recent OECD analyses*. OECD Publishing. Retrieved from http://www.oecd.org/site/educeri21st/40554299.pdf
Charlier, P., Cheron, M., Corbisier-Hagon, A. M., Dupont, C., Neven, M., Sénéca, G., & Wahl, J.-P. (2002). *Proposition de décret relative au pilotage du système éducatif de la Communauté française: Amendements déposés en commission de l'éducation* (Doc. 229 (2001-2002) n° 2, 28 February). Parlement de la Communauté Française.
Cheema, J. R. (2014). The migrant effect: An evaluation of native academic performance in Qatar. *Research in Education*, *91*(1), 65–77. http://doi.org/10.7227/RIE.91.1.6
Cheema, J. R. (2015). The private–public literacy divide amid educational reform in Qatar: What does PISA tell us? *International Review of Education*, *61*(2), 173–189. http://doi.org/10.1007/s11159-015-9479-8
Chen, H. T. (1990). *Theory-Driven Evaluations* (1st ed.). Thousand Oaks, CA: SAGE Publications.
Chenu, F., Crahay, M., & Lafontaine, D. (2014). Par-delà l'approche par compétences: quelle place réserver aux savoirs, à leur enseignement et à leur évaluation? *Education & Formation*, *e-302*, 17–29.
Cheron, M. (1995). *Proposition de décret définissant les objectifs généraux de l'enseignement* (Doc. 31 (SE 1995) n° 1, 20 September). Conseil de la Communauté Française.
Chingos, M. M. (2013). Class size and student outcomes: Research and policy implications. *Journal of Policy Analysis and Management*, *32*(2), 411–438. http://doi.org/10.1002/pam.21677
Chrostowski, S. J. (2004). Developing the TIMSS 2003 Background Questionnaires. In M. O. Martin, I. V. S. Mullis, & S. J. Chrostowski (Eds.), *TIMSS 2003 Technical Report* (pp. 66–91). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
Clerkin, A. (2013). Growth of the "Transition Year" programme nationally and in schools serving disadvantaged students, 1992–2011. *Irish Educational Studies*, *32*(2), 197–215. http://doi.org/10.1080/03323315.2013.770663
Coe, R., Aloisi, C., Higgins, S., & Elliot Major, L. (2014). *What makes great teaching? Review of the underpinning research*. The Sutton Trust. Retrieved from

http://www.suttontrust.com/researcharchive/great-teaching/
Cohen, A. (1979). Political Symbolism. *Annual Review of Anthropology*, *8*, 87–113. http://doi.org/10.1146/annurev.an.08.100179.000511
Coleman, J., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfeld, F., & York, R. (1966). *Equality of Educational Opportunity: Summary report* (Vol. 2). Washington, DC: US Department of Health, Education, and Welfare, Office of Education.
Coleman, J. S. (1988). Social Capital in the Creation of Human Capital. *American Journal of Sociology*, *94*, S95–S120. http://doi.org/10.2307/2780243
Commission de Pilotage. (2004a). *Avis à propos de l'organisation d'une épreuve commune en lien avec l'octroi du certificat d'étude de base* (Avis No. 1). Service général du Pilotage du système éducatif. Retrieved from http://www.enseignement.be/index.php?page=24771&navi=1010
Commission de Pilotage. (2004b). *PISA 2003 - Évaluation de la culture mathématique des jeunes de 15 ans: Document à l'attention des professeurs de mathématiques des 1er et 2ème degrés de l'enseignement secondaire*. Bruxelles: Service général du Pilotage du système éducatif. Retrieved from http://www.enseignement.be/download.php?do_id=2762&do_check=
Commission de Pilotage. (2006). *Proposition provisoire de la commission de pilotage relative aux améliorations a apporter a la formation initiale des enseignants afin de l'inscrire dans les objectifs du contrat pour l'école*. Service général du Pilotage du système éducatif. Retrieved from http://www.enseignement.be/index.php?page=24771&navi=1010
Commission de Pilotage. (2013). *Rapport intermédiaire de la Commission de Pilotage relatif au décret inscriptions*. Service général du Pilotage du système éducatif. Retrieved from http://www.enseignement.be/index.php?page=24771&navi=1010
Commission de Pilotage. (2014). *Rapport de la Commission de Pilotage relatif au décret inscriptions*. Service général du Pilotage du système éducatif. Retrieved from http://www.enseignement.be/index.php?page=24771&navi=1010
Committee on Integrating the Science of Early Childhood Development. (2000). *From Neurons to Neighborhoods*. (J. P. Shonkoff & D. A. Phillips, Eds.). Washington, DC: National Academy Press.
Comptroller and Auditor General. (2006). *Educational Disadvantage Initiatives in the Primary Sector*. Dublin: Department of Education and Science.
Condron, D. J. (2011). Egalitarianism and Educational Excellence: Compatible Goals for Affluent Societies? *Educational Researcher*, *40*(2), 47–55. http://doi.org/10.3102/0013189X11401021
Connell, R., & Dados, N. (2014). Where in the world does neoliberalism come from?: The market agenda in southern perspective. *Theory and Society*, *43*(2), 117–138. http://doi.org/10.1007/s11186-014-9212-9
Connelly, R., Sullivan, A., & Jerrim, J. (2014). *Primary and secondary education and poverty review*. London: Centre for Longitudinal Studies, Institute of Education, University of London.
Convention on the Organisation for Economic Co-operation and Development. (1960). 14 December 1960, Paris. Retrieved from https://www.oecd.org/general/conventionontheorganisationforeconomicco-operationanddevelopment.htm
Convention Secretariat. (1994). *Report on the National Education Convention*. (J. Coolahan, Ed.). Dublin: The National Education Convention Secretariat.
Conway, P. F., Murphy, R., Rath, A., & Hall, K. (2009). *Learning to teach and its implications for the continuum of teacher education: A nine-country cross-national study*. The Teaching Council. Retrieved from http://www.teachingcouncil.ie/research/commissioned-research.1555.html
Conway, P. F., & Sloane, F. C. (2005). *International Trends in Post-Primary Mathematics Education: Perspectives on Learning, Teaching and Assessment* (Research Report No.

5). Dublin: NCCA.
Coolahan, J. (2007). *A Review Paper on Thinking and Policies Relating to Teacher Education in Ireland*. The Teaching Council. Retrieved from http://www.teachingcouncil.ie/research/commissioned-research.1555.html
Corbett, C., Hill, C., & St. Rose, A. (2008). *Where the girls are: the facts about gender equity in education*. Washington, DC: AAUW Educational Foundation.
Cornet, J., & Dupriez, V. (2004). Du prescrit décrétal aux pratiques dans la classe: comment les enseignants perçoivent-ils et assument-ils les questions relatives à l'hétérogénéité des élèves? In M. Frenay & C. Maroy (Eds.), *L'école, six ans après le décret «missions»* (pp. 253–278). Louvain: Presses universitaires de Louvain.
Coryn, C. L. S., Noakes, L. A., Westine, C. D., & Schröter, D. C. (2011). A Systematic Review of Theory-Driven Evaluation Practice From 1990 to 2009. *American Journal of Evaluation*, *32*(2), 199–226. http://doi.org/10.1177/1098214010389321
Cosgrove, J. (2011). *Does Student Engagement Explain Performance on PISA? Comparisons of Response Patterns on the PISA Tests Over Time*. Dublin: Educational Research Centre. Retrieved from http://www.erc.ie/documents/engagement_and_performance_over_time.pdf
Cosgrove, J., & Cartwright, F. (2014). Changes in achievement on PISA : the case of Ireland and implications for international assessment practice. *Large-Scale Assessments in Education*, *2*(2), 1–17. http://doi.org/10.1186/2196-0739-2-2
Cosgrove, J., & Moran, G. (2011). *Taking the PISA 2009 Test in Ireland: Students' Response Patterns on the Print and Digital Assessments*. Dublin: Educational Research Centre.
Cosgrove, J., Shiel, G., Archer, P., & Perkins, R. (2010). *Comparisons of Performance in Ireland PISA 2000 to PISA 2009: A Preliminary Report to the Department of Education and Skills*. Dublin: Educational Research Centre.
Cosgrove, J., Shiel, G., Oldham, E., & Sofroniou, N. (2004). A survey of mathematics teachers in Ireland. *The Irish Journal of Education*, *35*, 20–44.
Cosgrove, J., Shiel, G., Sofroniou, N., Zastrutzki, S., & Shortt, F. (2004). *Education for life: The achievements of 15-year-olds in Ireland in the second cycle of PISA* (Summary Report). Dublin: Educational Research Centre.
Cosgrove, J., Sofroniou, N., Kelly, A., & Shiel, G. (2003). *A Teacher's Guide to the Reading Literacy Achievements of Irish 15-Year Olds*. Dublin: Educational Research Centre.
Cour d'Arbitrage. (2001). Arrêt n° 49/2001 du 18 avril 2001. *Moniteur Belge*. NUMAC: 2001021240. Retrieved from http://www.ejustice.just.fgov.be/doc/rech_f.htm
Cracolici, M. F., Cuffaro, M., & Nijkamp, P. (2010). The measurement of economic, social and environmental performance of countries: A novel approach. *Social Indicators Research*, *95*, 339–356. http://doi.org/10.1007/s11205-009-9464-3
Craft, J., & Howlett, M. (2013). The dual dynamics of policy advisory systems: The impact of externalization and politicization on policy advice. *Policy and Society*, *32*(3), 187–197. http://doi.org/10.1016/j.polsoc.2013.07.001
Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281–302. http://doi.org/10.1037/h0040957
Cunha, F., & Heckman, J. J. (2010). *Investing in Our Young People*. Retrieved from http://ideas.repec.org/p/nbr/nberwo/16201.html
Cunningham, C. (1963). Policy and Practice. *Public Administration*, *41*(3), 229–238. http://doi.org/10.1111/j.1467-9299.1963.tb01786.x
Cytermann, J.-R. (2010). Classements, indicateurs et politiques éducatives en France. *Revue Internationale D'éducation de Sèvres*, *54*, 83–92.
Dale, R. (1999). Specifying globalization effects on national policy: a focus on the mechanisms. *Journal of Education Policy*, *14*(1), 1–17. http://doi.org/10.1080/026809399286468
Dale, R. (2000). Globalization and education: Demonstrating a "common world educational culture" or locating a "globally structured educational agenda"? *Educational Theory*,

50(4), 427–448. http://doi.org/10.1111/j.1741-5446.2000.00427.x
Dale, R., & Robertson, S. L. (2007). *New arenas of global governance and international organisations: reflections and directions*. *Transformations of the State and Global Governance*. Bristol: Centre for Globalisation, Education and Societies, University of Bristol. Retrieved from http://susanleerobertson.com/publications/
Daly, A., & Thomas, L. (2008). *What influences vocational choices?* Ormskirk: Widening Participation Research Centre, Edge Hill University.
Daly, E. (2009). Religious freedom as a function of power relations: dubious claims on pluralism in the denominational schools debate. *Irish Educational Studies*, *28*(3), 235–251. http://doi.org/10.1080/03323310903335377
Darling-hammond, L. (2015). Can Value Added Add Value to Teacher Evaluation? *Educational Researcher*, *44*(2), 132–137. http://doi.org/10.3102/0013189X15575346
Darragh, J. (2007). Universal design for early childhood education: Ensuring access and equity for all. *Early Childhood Education Journal*, *35*(2), 167–171. http://doi.org/10.1007/s10643-007-0177-4
de Commer, B. (2003). Peut-on aider Rosetta sans assassiner Mozart ? *Le Sel*, (2), 5–6. Retrieved from http://www.sel-setca.org
de Commer, B. (2005). De PISA 2000 au Contrat pour l'école. *Courrier Hebdomadaire Du CRISP*. http://doi.org/10.3917/cris.1878.0005
de Olano, D. (2010). Gewinner, Verlierer und Exoten – PISA in sieben weiteren Staaten. In P. Knodel, K. Martens, D. de Olano, & M. Popp (Eds.), *Das PISA Echo* (pp. 251–299). Frankfurt am Main: Campus Verlag.
DeBoer, J. (2010). Why the Fireworks?: Theoretical Perspectives on the Explosion in International Assessments. In *The Impact of International Achievement Studies on National Education Policymaking* (pp. 293–326). http://doi.org/10.1108/S1479-3679(2010)0000013014
Delahais, T., & Toulemonde, J. (2012). Applying contribution analysis: Lessons from five years of practice. *Evaluation*, *18*(3), 281–293. http://doi.org/10.1177/1356389012450810
DeLuca, S., & Dayton, E. (2009). Switching Social Contexts: The Effects of Housing Mobility and School Choice Programs on Youth Outcomes. *Annual Review of Sociology*, *35*(1), 457–491. http://doi.org/10.1146/annurev-soc-070308-120032
Delvaux, B., & Maroy, C. (2009). Débat sur la régulation des inscriptions scolaires en Belgique francophone: où se situent les désaccords? *Les Cahiers de Recherche Du Girsef*, *68*.
Demeuse, M. (2000). La politique de discrimination positive en Communauté française de Belgique. Une méthode d'attribution de moyens supplémentaires basées sur des indicateurs objectifs. *Cahiers Du Service de Pédagogie Expérimentale*, *1-2*, 115–135.
Demeuse, M., Derobertmasure, A., Friant, N., Herremans, T., Monseur, C., Uyttendaele, S., & Verdale, N. (2007). *Étude exploratoire sur la mise en œuvre de nouvelles mesures visant à lutter contre les phénomènes de ségrégation scolaire et d'iniquité au sein du système éducatif de la Communauté française de Belgique* (Research report). Retrieved from http://www.enseignement.be/index.php?page=26044&id_fiche=5176&dummy=25981
Deming, D. (2009). Early Childhood Intervention and Life-Cycle Skill Formation: Evidence from Head Start. *American Economic Journal: Applied Economics*, *1*(3), 111–134.
Demonty, I., & Fagnant, A. (2006). PISA 2003: en moyenne, des résultats moyens mais trop de jeunes en situation critique. *TRACeS de ChanGements*, *74*, 1–3. Retrieved from http://changement-egalite.be/spip.php?article758&lang=fr
Department for Work and Pensions, & Department for Education. (2011). *A new approach to child poverty: Tackling the causes of disadvantage and transforming families' lives*. *Department for Work and Pensions*. Norwich: The Stationery Office.
Department of Education and Science. (1992). *Education for a changing world: Green paper for education*. Dublin: The Stationery Office.

Department of Education and Science. (1995). *Charting our education future: White Paper on education* (Launch Copy). Department of Education and Science. Retrieved from http://www.education.ie/en/Publications/Policy-Reports/Charting-Our-Education-Future-White-Paper-On-Education-Launch-Copy-1995-.pdf
Department of Education and Science. (1997). *International Adult Literacy Survey: Results for Ireland*. Dublin: Department of Education and Science.
Department of Education and Science. (2004). *A Brief Description of the Irish Education System*. Dublin: Communications Unit, Department of Education and Science.
Department of Education and Science. (2005). *DEIS (Delivering Equality Of Opportunity In Schools): An Action Plan for Educational Inclusion*. Dublin: Department of Education and Science.
Department of Education and Science. (2006). Transition Year Programmes – Guidelines for Schools. Retrieved from http://www.education.ie/en/Schools-Colleges/Information/Curriculum-and-Syllabus/Transition-Year-/ty_transition_year_school_guidelines.pdf
Department of Education and Science, & Educational Research Centre. (2005). Second National PISA Symposium. Retrieved from http://www.erc.ie/documents/p03symposiumprogramme.pdf
Department of Education and Science, & NCCA. (2002). *Junior Certificate Mathematics: Guidelines for Teachers* (Draft). Dublin: The Stationery Office.
Department of Education and Science, & NCCA. (2006). *Junior Certificate Science: Guidelines for Teachers*. Dublin: The Stationery Office.
Department of Education and Skills. (2011a). Lessons from Research on the Impact of DEIS: An information note for schools. Retrieved from http://www.education.ie/en/Schools-Colleges/Services/DEIS-Delivering-Equality-of-Opportunity-in-Schools-/DEIS-Supporting-Information/DEIS-Evaluation.html
Department of Education and Skills. (2011b). *Literacy and numeracy for learning and life: The National Strategy to Improve Literacy and Numeracy among Children and Young People 2011-2020*. Dublin: Department of Education and Skills.
Department of Education and Skills. (2012). *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country background report for Ireland*. Retrieved from http://www.oecd.org/education/school/oecdreviewonevaluationandassessmentframeworksforimprovingschooloutcomescountrybackgroundreports.htm
Department of Education and Skills. (2014). DEIS: Delivering Equality of Opportunity in Schools. Retrieved September 3, 2015, from http://www.education.ie/en/Schools-Colleges/Services/DEIS-Delivering-Equality-of-Opportunity-in-Schools-/
Department of Enterprise Trade and Employment. (2006). *Strategy for Science, Technology and Innovation 2006-2013*. Department of Enterprise, Trade and Employment. Retrieved from http://www.djei.ie/science/technology/sciencestrategy.htm
Department of Justice Equality and Law Reform. (2005). *Planning For Diversity: The National Action Plan Against Racism 2005-2008*. Dublin: Department of Justice, Equality and Law Reform.
Department of Social Protection. (2015). Back to School Clothing and Footwear Allowance. Retrieved November 25, 2015, from http://www.citizensinformation.ie/en/social_welfare/social_welfare_payments/social_welfare_payments_to_families_and_children/back_to_school_clothing_and_footwear_allowance.html
Desimone, L. M. (2009). Complementary Methods for Policy Research. In G. Sykes, B. Schneider, D. N. Plank, & T. G. Ford (Eds.), *Handbook of Educational Policy Research* (pp. 163–175). Oxon: Routledge.
Desrosières, A. (2002). *The Politics of Large Numbers: a History of Statistical Reasoning*. Cambridge, MA: Harvard University Press.
Destatte, P. (1995). Du rêve autonomiste à la souveraineté internationale. In F. Joris (Ed.),

Wallonie. Atouts et références d'une Région. Namur: Gouvernement wallon. Retrieved from http://www.wallonie-en-ligne.net/1995_Wallonie_Atouts-References/1995_ch03-1_Destatte_Philippe.htm
Dewey, J. (1915). *The School and Society*. Chicago, IL: University of Chicago Press.
Dickerson, A., McIntosh, S., & Valente, C. (2015). Do the Maths: An analysis of the gender gap in mathematics in Africa. *Economics of Education Review*, *46*, 1–22. http://doi.org/http://dx.doi.org/10.1016/j.econedurev.2015.02.005
Dillon, J. (2009). On scientific literacy and curriculum reform. *International Journal of Environmental and Science …*, *4*(3), 201–213. Retrieved from http://www.ijese.com/IJESE_Volume4_Issue3_July_2009.pdf#page=11
DiPrete, T. A., & Jennings, J. L. (2012). Social and behavioral skills and the gender gap in early educational achievement. *Social Science Research*, *41*(1), 1–15. http://doi.org/10.1016/j.ssresearch.2011.09.001
Dobbins, M., & Martens, K. (2012). Towards an education approach à la finlandaise? French education policy after PISA. *Journal of Education Policy*, *27*(1), 23–43. http://doi.org/10.1080/02680939.2011.622413
Dolin, J., & Krogh, L. B. (2010). The relevance and consequences of PISA science in a Danish context. *International Journal of Science and Mathematics Education*, *8*, 565–592. http://doi.org/10.1007/s10763-010-9207-6
Dolz, J., & Ollagnier, E. (Eds.). (2002). *L'énigme de la compétence en éducation*. Bruxelles: De Boeck. Retrieved from http://www.unige.ch/fapse/publications-ssed/RaisonsEducatives/REenligne/ECOED.html
Dorussen, H., Lenz, H., & Blavoukos, S. (2005). Assessing the Reliability and Validity of Expert Interviews. *European Union Politics*, *6*(3), 315–337. http://doi.org/10.1177/1465116505054835
Draelants, H. (2007). Les savoirs pédagogiques comme source de légimitation pour l'action publique en éducation. *Les Cahiers de Recherche Du Girsef*, *59*.
Draelants, H. (2014). Le choix de l'école en Belgique francophone: de l'individualisation à la bureaucratisation. *Les Cahiers de Recherche Du Girsef*, *99*.
Drudy, S. (2000). The Teaching Profession at a Time of Change: Vistas and Challenges for the Future. *Issues in Education*, *5*, 55–63.
DTI. (1998). *Our Competitive Future: Building the Knowledge Driven Economy*. London: Department of Trade and Industry.
Dudaite, J., & Elijio, A. (2008). *Trends in Similarities and Differences of Students' Mathematics Profiles in Various Countries* (3rd IEA International Research Conference, 18–20 September 2008). Taipei. Retrieved from http://www.iea.nl/irc-2008.html
Duncan, G. J., Ludwig, J., & Magnuson, K. A. (2010). Child Development. In P. B. Levine & D. J. Zimmerman (Eds.), *Targeting Investments in Children: Fighting Poverty When Resources are Limited* (pp. 27–58). Chicago, IL: University of Chicago Press.
Dunleavy, P., & Hood, C. (1994). From old public administration to new public management. *Public Money & Management*, *14*(3), 9–16. http://doi.org/10.1080/09540969409387823
Dupont, C., Neven, M., & Deghilaghe, F. (1997). *Rapport de commission présenté au nom de la Commission de l'Education et Rapport préliminaire, présenté par M. Deghilage, sur les auditions tenues à l'occasion du débat général sur l'enseignement obligatoire* (Doc. 152 (1996-1997) n° 62, 12 July). Parlement de la Communauté Française.
Dupriez, V., & Dumay, X. (2008). L'égalité à l'école: qu'en pensent les enseignants du primaire en Belgique francophone? *Revue Française de Pédagogie*, *163*, 77–89.
Duru-Bellat, M., & Suchaut, B. (2005). Organisation and Context, Efficiency and Equity of Educational Systems: what PISA tells us. *European Educational Research Journal*, *4*(3), 181. http://doi.org/10.2304/eerj.2005.4.3.3
Earl, L., Watson, N., & Torrance, N. (2002). Front row seats: what we've learned from the national literacy and numeracy strategies in England. *Journal of Educational Change*,

3, 35–53.
Education Endowment Foundation. (2016). Repeating a year. Retrieved March 2, 2016, from https://educationendowmentfoundation.org.uk/evidence/teaching-learning-toolkit/repeating-a-year/
Educational Disadvantage Committee. (2003a). *A More Integrated and Effective Delivery of School-based Educational Inclusion Measures*. Retrieved from www.into.ie/ROI/InfoforTeachers/SocialInclusion/EducationalDisadvantage/ReportsResources/EdDisSubmission2003.doc
Educational Disadvantage Committee. (2003b). *Educational Disadvantage Forum: Report of inaugural meeting*. Dublin: Educational Disadvantage Committee.
Educational Disadvantage Committee. (2005). *Moving beyond Educational Disadvantage: Report of the Educational Disadvantage Committee 2002-2005*. Dublin: Educational Disadvantage Committee.
EENEE. (2015). Andreas Schleicher of the OECD. Retrieved July 21, 2015, from http://www.eenee.de/eeneeHome/EENEE/Network-Members/Schleicher.html
Egalite, A. J. (2013a). Measuring Competitive Effects From School Voucher Programs: A Systematic Review. *Journal of School Choice*, *7*(4), 443–464. http://doi.org/10.1080/15582159.2013.837759
Egalite, A. J. (2013b). Response to John Merrifield's Comment, "Can We Reasonably Assess 'Productivity of Market-Based Educational Reforms'?" *Journal of School Choice*, *7*(4), 468–470. http://doi.org/10.1080/15582159.2013.837764
Egan, E. (2010). School improvement through external inspection and school self-evaluation. In *SDPI Summer School* (pp. 1–61). Galway. Retrieved from http://www.sdpi.ie/resources-site_map.html
Ehren, M. C. M., Shackleton, N., Tymms, P., Jones, K., Anane, E., Gustafsson, J., … Greger, D. (2014). *Impact of School Inspections on Teaching and Learning* (Technical Report). Retrieved from http://schoolinspections.eu/final-technical-report-entire-project-after-three-years/
Eichengreen, B., & Kenen, P. B. (1994). Managing the World Economy under the Bretton Woods System: An Overview. In P. B. Kenen (Ed.), *Managing the World Economy: fifty years after Bretton Woods* (pp. 3–87). Washington, DC: Institute for International Economics.
Eijkelhof, H. M. C., Kordes, J. H., & Savelsbergh, E. R. (2013). Implications of PISA Outcomes for Science Curriculum Reform in the Netherlands. In M. Prenzel, M. Kobarg, K. Schöps, & S. Rönnebeck (Eds.), *Research on PISA* (pp. 7–21). Dordrecht: Springer Science+Business Media B.V. http://doi.org/10.1007/978-94-007-4458-5_1
Eisenhardt, K. M. (1989). Agency Theory: An Assessment and Review. *The Academy of Management Review*, *14*(1), 57. http://doi.org/10.2307/258191
Eivers, E. (2010). PISA: issues in implementation and interpretation. *The Irish Journal of Education*, *38*, 94–118.
Eivers, E., Close, S., Shiel, G., Millar, D., Clerkin, A., Gilleece, L., & Kiniry, J. (2010). *The 2009 National Assessments of Mathematics and English Reading*. Dublin: Department of Education and Skills.
Eivers, E., & Creaven, A. (2013). Home-school interaction. In E. Eivers & A. Clerkin (Eds.), *National Schools, international contexts: Beyond the PIRLS and TIMSS test results* (pp. 105–128). Dublin: Educational Research Centre.
Eivers, E., Shiel, G., & Cheevers, C. (2006). *Implementing the Revised Junior Certificate Science Syllabus*. Dublin: The Stationery Office.
Eivers, E., Shiel, G., & Cunningham, R. (2007). *Ready for Tomorrow's World? The Competencies of Irish 15-year-olds in PISA 2006* (Summary Report). Dublin: The Stationery Office.
Eivers, E., Shiel, G., & Cunningham, R. (2008). *Ready for Tomorrow's World? The competencies of Ireland's 15-year-olds in PISA 2006* (Vol. Main repor). Dublin: Educational Research Centre.

Eivers, E., Shiel, G., & Shortt, F. (2004). *Reading literacy in disadvantaged primary schools*. Dublin: Educational Research Centre.
Eivers, E., Shiel, G., & Shortt, F. (2005). *Literacy in disadvantaged primary schools: Problems and solutions*. Dublin: Educational Research Centre.
Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin*, *136*(1), 103–127. http://doi.org/10.1037/a0018053
Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: a new look at an old issue. *Psychological Methods*, *12*(2), 121–38. http://doi.org/10.1037/1082-989X.12.2.121
Engle, P. L., Black, M. M., Behrman, J. R., Cabral de Mello, M., Gertler, P. J., Kapiriri, L., … Young, M. E. (2007). Strategies to avoid the loss of developmental potential in more than 200 million children in the developing world. *Lancet*, *369*, 229–242. http://doi.org/10.1016/S0140-6736(07)60112-3
Engle, P. L., Gamer, G. N., & Vargas-Barón, E. (2011). *Early childhood develoment: What Parliamentarians Need to Know and Do*. Geneva.
ERC. (2005). The Second National PISA Symposium. Dublin. Retrieved from http://www.erc.ie/documents/p03symposiumproceedings.pdf
ERC. (2015). How is PISA carried out? Retrieved August 28, 2015, from http://www.erc.ie/?p=58
Errity, D. (2014). Feedback from School Principals: January-March 2014. In *Learning from DEIS*. Dublin. Retrieved from http://www.education.ie/en/Schools-Colleges/Services/DEIS-Delivering-Equality-of-Opportunity-%E2%80%8Ein-Schools-/DEIS-Supporting-Information/Learning-from-DEIS-Seminar.html
ERT. (1995). *Education for Europeans: Towards the Learning Society*. Bruxelles: European Round Table of Industrialists.
Ertl, H. (2006). Educational standards and the changing discourse on education: the reception and consequences of the PISA study in Germany. *Oxford Review of Education*, *32*(5), 619–634. http://doi.org/10.1080/03054980600976320
ESRC. (2015). *ESRC Framework for Research Ethics*. Economic and Social Research Council. Retrieved from http://www.esrcsocietytoday.ac.uk/about-esrc/information/framework-for-research-ethics/index.aspx
Eurydice. (2001). *National actions to implement Lifelong Learning in Europe*. Brussels: Eurydice European Unit.
Eurydice. (2009a). *Early Childhood Education and Care in Europe: Tackling Social and Cultural Inequalities*. Brussels: EACEA P9 Eurydice. http://doi.org/10.2797/18055
Eurydice. (2009b). *Gender Differences in Educational Outcomes: Study on the Measures Taken and the Current Situation in Europe*. Brussels: EACEA P9 Eurydice.
Eurydice. (2016a). Belgium (French Comunity): Funding in Education. In *Eurypedia: European Encyclopedia on National Education Systems*. EACEA P9 Eurydice. Retrieved from https://webgate.ec.europa.eu/fpfis/mwikis/eurydice/index.php/Belgium-French-Community:Funding_in_Education
Eurydice. (2016b). Eurypedia: European Encyclopedia on National Education Systems. Retrieved from https://webgate.ec.europa.eu/fpfis/mwikis/eurydice/index.php?title=Home
Eurydice. (2016c). Ireland: Organisation and Governance. In *Eurypedia: European Encyclopedia on National Education Systems*. EACEA P9 Eurydice. Retrieved from https://webgate.ec.europa.eu/fpfis/mwikis/eurydice/index.php/Ireland:Organisation_and_Governance
Evans, D. K., & Kosec, K. (2012). *Early Child Education: Making Programs Work for Brazil's Most Important Generation*. Washington, DC: The World Bank. Retrieved from http://elibrary.worldbank.org/content/book/9780821389317
Exley, S. (2013). Mind the gap between best and worst: it's widening [TES magazine

online]. Retrieved November 16, 2015, from https://www.tes.com/article.aspx?storycode=6318807
Fairbrother, M. (2013). *Two Multilevel Modeling Techniques for Analyzing Comparative Longitudinal Survey Datasets* (Working Paper). Retrieved from http://seis.bris.ac.uk/~ggmhf/MHF.MLM-longit.2013.pdf
Fairclough, I., & Fairclough, N. (2012). *Political discourse analysis: A method for advanced students*. Oxon: Routledge.
Farías, M., & Sevilla, M. P. (2015). Effectiveness of Vocational High Schools in Students' Access to and Persistence in Postsecondary Vocational Education. *Research in Higher Education*. http://doi.org/10.1007/s11162-015-9370-2
Fenwick, L. (2011). Curriculum reform and reproducing inequality in upper-secondary education. *Journal of Curriculum Studies*, *43*(6), 697–716. http://doi.org/10.1080/00220272.2011.576773
Ferreira, J. A., Santos, E. J. R., Fonseca, A. C., & Haase, R. F. (2007). Early predictors of career development: A 10-year follow-up study. *Journal of Vocational Behavior*, *70*(1), 61–77. http://doi.org/10.1016/j.jvb.2006.04.006
Figazzolo, L. (2009). *Impact of PISA 2006 on the education policy debate*. Retrieved from http://download.ei-ie.org/docs/IRISDocuments/Research Website Documents/2009-00036-01-E.pdf
Figlio, D. N., & Hart, C. M. D. (2014). Competitive effects of means-tested school vouchers. *American Economic Journal: Applied Economics*, *6*(1), 133–156. http://doi.org/10.1257/app.6.1.133
Firth, R. (1973). *Symbols: Public and Private*. Ithaca, NY: Cornell University Press.
Fischbach, A., Keller, U., Preckel, F., & Brunner, M. (2013). PISA proficiency scores predict educational outcomes. *Learning and Individual Differences*, *24*, 63–72. http://doi.org/10.1016/j.lindif.2012.10.012
Fitzgerald, E. (2001). Redistribution through Ireland's Welfare and Tax Systems. In S. Cantillon, C. Corrigan, P. Kirby, & J. O'Flynn (Eds.), *Rich and Poor: Perspectives on tackling inequality in Ireland*. Dublin: Oak Tree Press.
Fitz-Gibbon, C. T., & Morris, L. L. (1996). Theory-based evaluation. *Evaluation Practice*, *17*(2), 177–184. http://doi.org/10.1016/S0886-1633(96)90024-0
Fleming, C. B., Harachi, T. W., Catalano, R. F., Haggerty, K. P., & Abbott, R. D. (2001). *Assessing the effects of a school-based intervention on unscheduled school transfers during elementary school*. *Evaluation review* (Vol. 25).
Forfás. (2004). *Forfás Annual Report 2003*. Dublin: Forfás.
Freeman, J., & Simonsen, B. (2015). Examining the Impact of Policy and Practice Interventions on High School Dropout and School Completion Rates: A Systematic Review of the Literature. *Review of Educational Research*, *85*(2), 205 –248. http://doi.org/10.3102/0034654314554431
Friant, N., Demeuse, M., Nicaise, I., & Aubert-lotarski, A. (2008). Les politiques d'éducation prioritaire en Belgique: deux modes de régulation des effets d'une logique de marché. In *Évolution des politiques d'éducation prioritaire en Europe: Conceptions, mises en œuvres, débats* (pp. 85–134). Lyon: Institut National de Recherche Pédagogique.
Friant, N., Derobertmasure, A., & Demeuse, M. (2008). Les liens entre recherche prospective, description du système éducatif et pilotage: un exemple en Communauté française de Belgique. In *Entre la régulation des apprentissages et le pilotage des systèmes: évaluations en tension. Actes du 20e colloque de l'ADMEE-Europe*. Retrieved from https://plone.unige.ch/sites/admee08/communications-individuelles/v-a4/v-a4-2
Friedman Foundation for Educational Choice. (2015). How School Vouchers Work. Retrieved December 9, 2015, from http://www.edchoice.org/school-choice/what-is-school-choice/
Fuchs, T., & Wößmann, L. (2007). What accounts for international differences in student

performance? A re-examination using PISA data. *Empirical Economics*, *32*(2-3), 433–464. http://doi.org/10.1007/s00181-006-0087-0
Fullan, M. (2000). The return of large-scale reform. *Journal of Educational Change*, *1*, 5–28.
Fullan, M. (2007). *The New Meaning of Educational Change* (4th ed.). New York, NY: Teachers College Press.
Fullan, M. (2009). Large-scale reform comes of age. *Journal of Educational Change*, *10*(2-3), 101–113. http://doi.org/10.1007/s10833-009-9108-z
Funnell, S. C. (1997). Program Logic: An Adaptable Tool for Designing and Evaluating Programs. *Evaluation News and Comment*, *6*(1), 5–7.
Garrouste, C. (2010). *100 Years of Educational Reforms in Europe: a contextual database*. Luxembourg: Publications Office of the European Union.
Garvin, T. (2004). *Preventing the Future: Why Was Ireland So Poor for So Long?* Dublin: Gill & Macmillan.
Gebhardt, E., & Adams, R. J. (2007). The influence of equating methodology on reported trends in PISA. *Journal of Applied Measurement*, *8*(3), 305–322. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/17804896
Gelman, A. (2013). Commentary: P Values and Statistical Practice. *Epidemiology*, *24*(1), 69–72. http://doi.org/10.1097/EDE.0b013e31827886f7
Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. New York, NY: Cambridge University Press.
Gersten, R., Dimino, J., Jayanthi, M., Kim, J. S., & Santoro, L. E. (2010). Teacher Study Group: Impact of the Professional Development Model on Reading Instruction and Student Outcomes in First Grade Classrooms. *American Educational Research Journal*, *47*(3), 694–739. http://doi.org/10.3102/0002831209361208
Gersten, R., Taylor, M. J., Keys, T. D., Rolfhus, E., & Newman-gonchar, R. (2014). *Summary of research on the effectiveness of math professional development approaches*. Washington, DC: National Center for Education Evaluation and Regional Assistance.
Gerth, H. H., & Wright Mills, C. (1946). *From Max Weber: Essays in Sociology*. New York: Oxford University Press.
Gibbons, S., & Mcnally, S. (2013). *The Effects of Resources Across School Phases: A Summary of Recent Evidence* (CEP Discussion Paper No. 1226). London: Centre for Economic Performance, London School of Economics and Political Science.
Gilmore, A. (2005). *The impact of PIRLS (2001) and TIMSS (2003) in low- and middle-income countries*. Retrieved from http://www.iea.nl/fileadmin/user_upload/Publications/Electronic_versions/Gilmore_Impact_PIRLS_TIMSS.pdf
Goldstein, H. (2004). International comparisons of student attainment: some issues arising from the PISA study. *Assessment in Education: Principles, Policy & Practice*, *11*(3), 319–330. http://doi.org/10.1080/0969594042000304618
Goldstein, H. (2011). *Multilevel statistical models* (4th ed.). Chichester: John Wiley & Sons, Ltd.
Goodman, R. (2003). The why, what and how of educational reform in Japan. In R. Goodman & D. Philips (Eds.), *Can the Japanese Change Their Education System?* (pp. 7–30). Oxford: Symposium Books.
Goos, M., Schreier, B. M., Knipprath, H. M. E., De Fraine, B., Van Damme, J., & Trautwein, U. (2013). How Can Cross-Country Differences in the Practice of Grade Retention Be Explained? A Closer Look at National Educational Policy Factors. *Comparative Education Review*, *57*(1), 54–84. http://doi.org/10.1086/667655
Gormley, W. T. J., Gayer, T., Phillips, D., & Dawson, B. (2005). The effects of universal pre-K on cognitive development. *Developmental Psychology*, *41*(6), 872–884. http://doi.org/10.1037/0012-1649.41.6.872
Gouvernement de la Communauté française. (2004). *Déclaration commune entre le*

Gouvernement de la Communauté française, les Organisations représentatives de la Communauté éducative et les Partenaires sociaux: Objectifs, principes généraux et priorités du Contrat Strategique pour l'Education. Bruxelles. Retrieved from http://www.contrateducation.be/contenu_telechargements.asp

Gouvernement de la Communauté française. (2005a). *Contrat pour l'école*. Retrieved from http://www.contrateducation.be/contenu_telechargements.asp

Gouvernement de la Communauté française. (2005b). *Projet de Contrat stratégique pour l'Education*. Retrieved from http://www.contrateducation.be/contenu_telechargements.asp

Gouvernement de la Communauté française. (2009). *Évaluation du Countrat pour l'École*. Bruxelles: Communauté Française de Belgique. Retrieved from http://cms.horus.be/files/99940/MediaArchive/EvaluationCE260309.pdf

Gouvernement de la Communauté française. (2014). *Fédérer pour Réussir: Déclaration de politique communautaire 2014-2019*. Retrieved from http://www.federation-wallonie-bruxelles.be/index.php?id=dpc2014-2019

Gove, M. (2012). Education for economic success. Speech to the Education World Forum, 11 January 2012, London. Retrieved from https://www.gov.uk/government/speeches/michael-gove-to-the-education-world-forum

Government of Ireland. (1965). *Investment in education*. *Educational Researcher*. Dublin: The Stationery Office.

Government of Ireland. (1997). *National Anti-Poverty Strategy*. Dublin: The Stationery Office.

Government of Ireland. (2000). *Ireland: National Development Plan 2000-2006*. Dublin: The Stationery Office.

Government of Ireland. (2007). *Transforming Ireland: A Better Quality of Life for All* (National Development Plan 2007-2013). Dublin: The Stationery Office.

Grannell, J. J., Barry, P. D., Cronin, M., Holland, F., & Hurley, D. (2011). *Interim report on Project Maths*. Retrieved from http://www.ucc.ie/en/media/academic/maths/ProjectMathsInterimReport_Nov2011.pdf

Grek, S. (2009). Governing by numbers: the PISA "effect" in Europe. *Journal of Education Policy*, *24*(1), 23–37. http://doi.org/10.1080/02680930802412669

Grek, S. (2012). What PISA knows and can do: Studying the role of national actors in the making of PISA. *European Educational Research Journal*, *11*(2), 243–254. http://doi.org/10.2304/eerj.2012.11.2.243

Grek, S., Lawn, M., Lingard, B., Ozga, J., Rinne, R., Segerholm, C., & Simola, H. (2009). National policy brokering and the construction of the European Education Space in England, Sweden, Finland and Scotland. *Comparative Education*, *45*(1), 5–21. http://doi.org/10.1080/03050060802661378

Grisay, A., De Jong, J. H. A. L., Gebhardt, E., Berezner, A., & Halleux-Monseur, B. (2007). Translation equivalence across PISA countries. *Journal of Applied Measurement*, *8*(3), 249–266.

Grisay, A., Gonzalez, E. J., & Monseur, C. (2009). Equivalence of item difficulties across national versions of the PIRLS and PISA reading assessments. In M. von Davier & D. Hastedt (Eds.), *IERI monograph series: Issues and methodologies in large-scale assessments* (Vol. 2, pp. 63–84). IEA-ETS Research Institute (IERI).

Grootaers, D. (2005). Les mutations de l'égalité des chances à l'école. *Courrier Hebdomadaire Du CRISP*, (28).

Grootaers, D. (2014). Le tronc commun dans l'enseignement secondaire. *Courrier Hebdomadaire Du CRISP*, (5).

Gruman, D. H., Harachi, T. W., Abbott, R. D., Catalano, R. F., & Fleming, C. B. (2008). Longitudinal Effects of Student Mobility on Three Dimensions of Elementary School Engagement. *Child Development*, *79*(6), 1833–1852. http://doi.org/10.1111/j.1467-8624.2008.01229.x

Guarino, C. M., & Tanner, J. C. (2012). Adequacy, accountability, autonomy and equity in a

Middle Eastern school reform: The case of Qatar. *International Review of Education*, *58*(2), 221–245. http://doi.org/10.1007/s11159-012-9286-4
Gugiu, P. C., & Rodríguez-Campos, L. (2007). Semi-structured interview protocol for constructing logic models. *Evaluation and Program Planning*, *30*(4), 339–50. http://doi.org/10.1016/j.evalprogplan.2007.08.004
Guidet, P., & Bouillon, P. (1990). Les enseignants déchaînés étouffent la fête à Namur: Sept bougies sur le gâteau de la fronde scolaire. *Le Soir*. 27 September 1990. Retrieved from http://archives.lesoir.be/les-enseignants-dechaines-etouffent-la-fete-a-namur-sep_t-19900927-Z034F2.html
Gür, B. S., Çelik, Z., & Özoğlu, M. (2011). Policy options for Turkey: a critique of the interpretation and utilization of PISA results in Turkey. *Journal of Education Policy*, *27*(1), 1–21. http://doi.org/10.1080/02680939.2011.595509
Haahr, J. H., Nielsen, T. K., Hansen, M. E., & Jakobsen, S. T. (2005). *Explaining Student Performance: Evidence from the international PISA, TIMSS and PIRLS surveys*. Retrieved from https://www.oecd.org/education/school/programmeforinternationalstudentassessmentpisa/35920726.pdf
Haas, P. M. (1992). *Introduction: epistemic communities and international policy coordination*. *International Organization* (Vol. 46). http://doi.org/10.1017/S0020818300001442
Haertel, E. H. (2004). *The Behavior of Linking Items in Test Equating* (CSE Report No. 630). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing (CRESST). Retrieved from http://www.cse.ucla.edu/products/Reports/R630.pdf
Hallinger, P., Heck, R. H., & Murphy, J. (2014). Teacher evaluation and school improvement: An analysis of the evidence. *Educational Assessment, Evaluation and Accountability*, *26*(1), 5–28. http://doi.org/10.1007/s11092-013-9179-5
Halpin, D., & Troyna, B. (1995). The Politics of Education Policy Borrowing. *Comparative Education*, *31*(3), 303–310. http://doi.org/10.1080/03050069528994
Halton, M. J. (2003). Benchmarking: another attempt to introduce market-oriented policies into Irish second-level education? *Pedagogy, Culture & Society*, *11*(3), 331–352.
Hanberger, A. (2014). What PISA intends to and can possibly achieve: A critical programme theory analysis. *European Educational Research Journal*, *13*(2), 167–180. http://doi.org/10.2304/eerj.2014.13.2.167
Hanushek, E. A., Link, S., & Wößmann, L. (2013). Does school autonomy make sense everywhere? Panel estimates from PISA. *Journal of Development Economics*, *104*, 212–232. http://doi.org/10.1016/j.jdeveco.2012.08.002
Hanushek, E. A., & Rivkin, S. G. (2006). Teacher Quality. In E. A. Hanushek & F. Welch (Eds.), *Handbook of the Economics of Education* (Vol. 2, pp. 1051–1078). Amsterdam: North Holland. http://doi.org/10.1016/S1574-0692(06)02018-6
Hanushek, E. A., & Rivkin, S. G. (2012). The Distribution of Teacher Quality and Implications for Policy. *Annual Review of Economics*, *4*(1), 131–157. http://doi.org/10.1146/annurev-economics-080511-111001
Hanushek, E. A., & Wößmann, L. (2011). The Economics of International Differences in Educational Achievement. In E. A. Hanushek, S. Machin, & L. Wößmann (Eds.), *Handbook of the Economics of Education* (Vol. 3, pp. 89–200). Amsterdam: North-Holland. http://doi.org/10.1016/B978-0-444-53429-3.00002-8
Hanushek, E. A., & Wößmann, L. (2012). Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation. *Journal of Economic Growth*, *17*(4), 267–321. http://doi.org/10.1007/s10887-012-9081-x
Hargreaves, A. (2009). The fourth way of change. In *Change Wars* (pp. 11–44). Bloomington, IN: Solution Tree.
Hargreaves, A., & Fink, D. (2006). *Sustainable leadership*. San Francisco, CA: Jossey-Bass.
Harris, A. (2011). Reforming systems: Realizing the Fourth Way. *Journal of Educational*

Change, *12*(2), 159–171. http://doi.org/10.1007/s10833-011-9156-z
Hartley, B. L., & Sutton, R. M. (2013). A Stereotype Threat Account of Boys' Academic Underachievement. *Child Development*, *84*(5), 1716–1733. http://doi.org/10.1111/cdev.12079
Hartong, S. (2012). Overcoming resistance to change: PISA, school reform in Germany and the example of Lower Saxony. *Journal of Education Policy*, *27*(6), 747–760. http://doi.org/10.1080/02680939.2012.672657
Harvey-Beavis, A. (2002). Student and School Questionnaire Development. In R. J. Adams & M. L. Wu (Eds.), *PISA 2000 Technical Report* (pp. 33–38). Paris: OECD Publishing.
Hattie, J. (2008). *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement*. New York, NY: Routledge.
Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, *77*(1), 81–112. http://doi.org/10.3102/003465430298487
Hawkins, D. G., Lake, D. A., Nielson, D. L., & Tierney, M. J. (Eds.). (2006). *Delegation and Agency in International Organizations*. New York, NY: Cambridge University Press.
Hayden, F. G. (2006). *Policymaking for A Good Society: The Social Fabric Matrix Approach to Policy Analysis and Program Evaluation*. New York, NY: Springer Science+Business Media, Inc.
He, Q., & Tymms, P. (2013). The principal axis approach to value-added calculation. *Educational Research and Evaluation*, 1–19. http://doi.org/10.1080/13803611.2013.860036
Heck, R. H. (2004). *Studying Educational and Social Policy: Theoretical Concepts and Research Methods*. Mahwah, NJ: Lawrence Erlbaum Associates.
Heckman, J. J. (2008). Schools, skills, and synapses. *Economic Inquiry*, *46*(3), 289–324. http://doi.org/10.1111/j.1465-7295.2008.00163.x
Heichel, S., Holzinger, K., Sommerer, T., Liefferink, D., Pape, J., & Veenman, S. (2008). Research design, variables and data. In K. Holzinger, C. Knill, & B. Arts (Eds.), *Environmental Policy Convergence in Europe: The Impact of International Institutions and Trade* (pp. 64–97). New York, NY: Cambridge University Press.
Helbach, C. (2012). *The Interplay of Standardized Tests and Incentives: An Econometric Analysis with Data from PISA 2000 and PISA 2009* (Ruhr Economic Papers No. 356).
Heller, J. I., Daehler, K. R., Wong, N., Shinohara, M., & Miratrix, L. W. (2012). Differential effects of three professional development models on teacher knowledge and student achievement in elementary science. *Journal of Research in Science Teaching*, *49*(3), 333–362. http://doi.org/10.1002/tea.21004
Henry, M., Lingard, B., Rizvi, F., & Taylor, S. (2001). *The OECD, globalisation, and education policy*. Amsterdam: Pergamon for the International Association of Universities.
Hinchey, P. H. (2010). *Getting Teacher Assessment Right: What Policymakers Can Learn From Research*. *National Education Policy Center*. Boulder, CO: National Education Policy Center. Retrieved from http://nepc.colorado.edu/files/PB-TEval-Hinchey_0.pdf
Hindriks, J., Verschelde, M., Rayp, G., & Schoors, K. (2010). *School autonomy and educational performance: within-country evidence* (CORE discussion paper No. 82). Retrieved from http://econpapers.repec.org/paper/corlouvco/2010082.htm
Hinnant, J. B., O'Brien, M., & Ghazarian, S. R. (2009). The Longitudinal Relations of Teacher Expectations to Achievement in the Early School Years. *Journal of Educational Psychology*, *101*(3), 662–670. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2860190/
Hodgson, A., & Spours, K. (2010). Vocational qualifications and progression to higher education: the case of the 14–19 Diplomas in the English system. *Journal of Education and Work*, *23*(2), 95–110. http://doi.org/10.1080/13639080903578619
Hoelscher, M., Hayward, G., Ertl, H., & Dunbar- Goddet, H. (2008). The transition from vocational education and training to higher education: a successful pathway? *Research*

Papers in Education, *23*(2), 139–151. http://doi.org/10.1080/02671520802048679
Hogenboom, J.-P. (2003). Fondements d'une didactique des compétences. *Informations Pédagogiques*, *54*, 11–19.
Holland, S. (1979). *Rutland Street: The story of an educational experiment for disadvantaged children in Dublin*. Oxford: Pergamon Press.
Holme, J. J., Richards, M. P., Jimerson, J. B., & Cohen, R. W. (2010). Assessing the Effects of High School Exit Examinations. *Review of Educational Research*, *80*(4), 476–526. http://doi.org/10.3102/0034654310383147
Holmes, C. T. (1989). Grade level retention effects: A meta-analysis of research studies. In L. A. Shepard & M. L. Smith (Eds.), (pp. 16–33). London: Falmer Press.
Holzinger, K., Knill, C., & Arts, B. (Eds.). (2008). *Environmental Policy Convergence in Europe: The Impact of International Institutions and Trade*. New York, NY: Cambridge University Press.
Honig, M. I. (2004). The New Middle Management: Intermediary Organizations in Education Policy Implementation. *Educational Evaluation and Policy Analysis*, *26*(1), 65–87. http://doi.org/10.3102/01623737026001065
Hood, C. (1991). A public management for all seasons? *Public Administration*, *69*(1), 3–19. http://doi.org/10.1111/j.1467-9299.1991.tb00779.x
Hopkins, D., Pennock, D., & Ritzen, J. (2008). *External evaluation of the policy impact of PISA* (26th meeting of the PISA Governing Board, 3-5 November). The Hague: OECD Publishing. Retrieved from http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=EDU/PISA/GB/M(2008)2/REV1&docLanguage=En
Hopkins, D., Stringfield, S., Harris, A., Stoll, L., & Mackay, T. (2014). School and system improvement: a narrative state-of-the-art review. *School Effectiveness and School Improvement*, *25*(2), 257–281. http://doi.org/10.1080/09243453.2014.885452
Hopmann, S. T., & Brinek, G. (2007). Introduction: PISA According to PISA – Does PISA Keep What It Promises? In S. T. Hopmann, G. Brinek, & M. Retzl (Eds.), *PISA zufolge PISA - PISA According to PISA* (pp. 10–19). Wien: Lit Verlag.
Hopmann, S. T., Brinek, G., & Retzl, M. (Eds.). (2007). *PISA zufolge PISA - PISA According to PISA*. Wien: Lit Verlag.
Horn, D. (2009). Age of selection counts: a cross-country analysis of educational institutions. *Educational Research and Evaluation*, *15*(4), 343–366. http://doi.org/10.1080/13803610903087011
Horn, D. (2013). Diverging performances: the detrimental effects of early educational selection on equality of opportunity in Hungary. *Research in Social Stratification and Mobility*, *32*, 25–43. http://doi.org/10.1016/j.rssm.2013.01.002
Houghton, J., & Sheehan, P. (2000). *A Primer on the Knowledge Economy* (CSES Working Paper No. 18). Melbourne: Centre for Strategic Economic Studies, Victoria University of Technology. Retrieved from http://www.cfses.com/documents/knowledgeeconprimer.pdf
Howlett, M., & Joshi-Koop, S. (2011). Transnational learning, policy analytical capacity, and environmental policy convergence: Survey results from Canada. *Global Environmental Change*, *21*, 85–92. http://doi.org/10.1016/j.gloenvcha.2010.10.002
Hoxby, C. M. (2003). School choice and school competition: Evidence from the United States. *Swedish Economic Policy Review*, *10*, 11–67.
Hubley, A. M., & Zumbo, B. D. (2011). Validity and the Consequences of Test Interpretation and Use. *Social Indicators Research*, *103*(2), 219–230. http://doi.org/10.1007/s11205-011-9843-4
Hustedt, J. T., Barnett, W. S., Jung, K., & Goetze, L. D. (2009). *The New Mexico PreK Evaluation: Results from the Initial Four Years of a New State Preschool Initiative* (Final report). New Brunswick, NJ: National Institute for Early Education Research, Rutgers University.
Hyland, Á. (2014). The investment in education report 1965 – recollections and

reminiscences. *Irish Educational Studies*, *33*(2), 123–139. http://doi.org/10.1080/03323315.2014.918297
IEA. (2010). About us. Retrieved from http://www.iea.nl/about_us.html
Inspectorate. (2003). *Looking at our School: An aid to self-evaluation in second-level schools*. Dublin: The Stationery Office.
Inspectorate. (2005). *Literacy and Numeracy in Disadvantaged Schools: Challenges for Teachers and Learners*. Dublin: The Stationery Office.
Inspectorate. (2006). *Looking at English: Teaching & learning English in post-primary schools*. Department of Education and Science. Retrieved from http://www.education.ie/en/Publications/Inspection-Reports-Publications/Evaluation-Reports-Guidelines/insp_looking_at_english_pdf.pdf
Inspectorate. (2008). *Looking at Junior Cycle Science*. Dublin: Department of Education and Science.
Inspectorate. (2010). *A Guide to Whole-School Evaluation in Primary Schools* (Revised June 2010). Dublin: Department of Education and Skills.
Inspectorate. (2011a). *An evaluation of planning processes in DEIS post-primary schools*. Dublin: Department of Education and Skills.
Inspectorate. (2011b). *An Evaluation of Planning Processes in DEIS Primary Schools*. Dublin: Department of Education and Skills.
Inspectorate. (2013). *Chief Inspector's Report 2010-2012*. Dublin: Inspectorate of the Department of Education and Skills.
Inspectorate. (2015). Inspection Reports and Publications: Forms of Inspection. Retrieved September 3, 2015, from http://www.education.ie/en/Publications/Inspection-Reports-Publications/
INTO. (1997). *Teaching & Assessment: Issues in assessment*. Dublin: Irish National Teachers' Organisation.
INTO. (2008a). Assessment in the Primary School. In *Consultative Conference on Education*. Dublin: Irish National Teachers' Organisation.
INTO. (2008b). *The Primary School Curriculum: INTO Survey 2005 and Proceedings of the Consultative Conference on Education 2006*. Dublin: Irish National Teachers' Organisation.
INTO Education Committee. (2004). *Change in Primary Education: Challenges for Teachers*. Dublin: Irish National Teachers' Organisation.
INTO Education Committee. (2009). *Whole School Evaluation*. Dublin: Irish National Teachers' Organisation.
Irish Maths Teachers Association. (2013). *Project Maths and the Irish Maths Teachers Association*. Retrieved from http://www.imta.ie/IMTA PM_1 Doc.pdf
Jakobi, A. P., & Teltemann, J. (2009). *Convergence and divergence in welfare state development: an assessment of education policy in OECD countries* (No. 93). *TranState working papers*. Retrieved from http://hdl.handle.net/10419/27915
Jakubowski, M. (2013). *Analysis of the Predictive Power of PISA Test Items* (OECD Education Working Papers No. 87). Retrieved from http://dx.doi.org/10.1787/5k4bx47268g5-en OECD
Jeffes, J., Jones, E., Cunningham, R., Dawson, A., Cooper, L., Straw, S., … Kane, M. O. (2012). *Research into the impact of Project Maths on student achievement, learning and motivation* (First Interim Report). Slough: National Foundation for Educational Research.
Jeffes, J., Jones, E., Wilson, M., Lamont, E., Straw, S., Wheater, R., & Dawson, A. (2013). *Research into the impact of Project Maths on student achievement, learning and motivation* (Final Report). Slough: National Foundation for Educational Research.
Jensen, B., Weidmann, B., & Farmer, J. (2013). *The myth of markets in school education*. Grattan Institute. Retrieved from http://grattan.edu.au/static/files/assets/de60db0d/myth_of_markets_in_school_education.pdf

Jerrim, J. (2014). The Unrealistic Educational Expectations of High School Pupils: Is America Exceptional? *The Sociological Quarterly*, *55*(1), 196–231. http://doi.org/10.1111/tsq.12049
Jiang, Z. (2014). Emotional Intelligence and Career Decision-Making Self-Efficacy: National and Gender Differences. *Journal of Employment Counseling*, *51*(3), 112–124. http://doi.org/10.1002/j.2161-1920.2014.00046.x
Joint Committee on Education and Skills. (2011). Ireland's Ranking on PISA: Discussion. *Joint Committee on Education and Skills Debate*. 13 January 2011. Retrieved from http://debates.oireachtas.ie/EDJ/2011/01/13/00004.asp
Jones, P. W. (2007). Education and World Order. *Comparative Education*, *43*(3), 325–337. http://doi.org/10.1080/03050060701556273
Jonnaert, P., Masciotra, D., Barrette, J., Morel, D., & Mane, Y. (2007). From Competence in the Curriculum To Competence in Action. *Prospects*, *37*(2), 187–203.
Joyce, B. R., & Showers, B. (2002). *Student achievement through staff development* (3rd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.
Jürges, H., Schneider, K., Senkbeil, M., & Carstensen, C. H. (2012). Assessment drives learning: The effect of central exit exams on curricular knowledge and mathematical literacy. *Economics of Education Review*, *31*(1), 56–65. http://doi.org/10.1016/j.econedurev.2011.08.007
Kane, M. T. (1992). An Argument-Based Approach to Validity. *Psychological Bulletin*, *112*(3), 527–535.
Kane, M. T. (2001). Current Concerns in Validity Theory. *Journal of Educational Measurement*, *38*(4), 319–342. http://doi.org/10.1111/j.1745-3984.2001.tb01130.x
Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), (4th ed., pp. 17–64). Westport, CT: American Council of Education/Praeger.
Kane, M. T. (2013a). Validating the Interpretations and Uses of Test Scores. *Journal of Educational Measurement*, *50*(1), 1–73. http://doi.org/10.1111/jedm.12000
Kane, M. T. (2013b). Validation as a Pragmatic, Scientific Activity. *Journal of Educational Measurement*, *50*(1), 115–122. http://doi.org/10.1111/jedm.12007
Kane, T. J., McCaffrey, D. F., & Staiger, D. O. (2013). *Have We Identified Effective Teachers?* MET Projet / Bill & Melinda Gates Foundation.
Kaufmann, D., Kraay, A., & Mastruzzi, M. (2010). *The worldwide governance indicators: methodology and analytical issues* (No. 5430). *Policy Research Working Paper*. The World Bank.
Kautz, T., Heckman, J. J., Diris, R., ter Weel, B., & Borghans, L. (2014). *Fostering and Measuring Skills: Improving Cognitive and Non-Cognitive Skills to Promote Lifetime Success*. Paris: OECD Publishing.
Kellaghan, T. (1977). *The evaluation of an intervention programme for disadvantaged children*. Windsor: National Foundation for Educational Research.
Kellaghan, T., McGee, P., Millar, D., & Perkins, R. (2004). *Views of the Irish public on education: 2004 survey*. Dublin: Educational Research Centre.
Kellaghan, T., Weir, S., Ó hUallacháin, S., & Morgan, M. (1995). *Educational Disadvantage In Ireland*. Dublin: Department of Education, Combat Poverty Agency, Educational Research Centre.
Kelley, T. L. (1927). *Interpretation of educational measurements*. World Book Company.
Kellow, A. J., & Carroll, P. G. H. (2013). Exploring the Impact of International Civil Servants: The Case of the OECD. *International Journal of Public Administration*, *36*(7), 482–491. http://doi.org/10.1080/01900692.2013.772635
Kelly, D. L. (2003). Developing the PIRLS Background Questionnaires. In M. O. Martin, I. V. S. Mullis, & A. M. Kennedy (Eds.), *PIRLS 2001 Technical Report* (pp. 29–40). Chestnut Hill, MA: International Study Center, Lynch School of Education, Boston College.
Kerstan, T. (2000, November 23). Thema verfehlt, Sechs! *Die Zeit*, (48). Retrieved from http://www.zeit.de/2000/48/Thema_verfehlt_Sechs_

Khandker, S. R., Koolwal, G. B., & Samad, H. (2009). *Handbook on Impact Evaluation: Quantitative Methods and Practices*. *Learning*. Washington, DC: The World Bank.
Kilkelly, U. (2007). *Barriers to the Realisation of Children's Rights in Ireland*. Dublin: Ombudsman for Children. Retrieved from http://ssrn.com/abstract=1961731
Kind, P. M. (2013). Establishing assessment scales using a novel disciplinary rationale for scientific reasoning. *Journal of Research in Science Teaching*, *50*(5), 530–560. http://doi.org/10.1002/tea.21086
Kingdon, J. W. (1995). *Agendas, Alternatives and Public Policies* (2nd ed.). New York, NY: HarperCollins.
Kirsch, I. (2001). *The international adult literacy survey (IALS): Understanding what was measured* (Research report No. RR-01-25). Princeton, NJ: Educational Testing Service.
Klemencic, E. (2010). The impact of international achievement studies on national education policymaking: The case of Slovenia – how many watches do we need? In *The Impact of International Achievement Studies on National Education Policymaking* (pp. 239–266). Elsevier. http://doi.org/10.1108/S1479-3679(2010)13
Klieme, E., & Kuger, S. (2014). *PISA 2015: Draft questionnaire framework*. Retrieved from https://www.oecd.org/pisa/pisaproducts/
Knighton, T., & Bussière, P. (2006). *Educational Outcomes at Age 19 Associated with Reading Ability at Age 15*. Ottawa: Statistics Canada. Retrieved from http://publications.gc.ca/collections/Collection/Statcan/81-595-MIE/81-595-MIE2006043.pdf
Knodel, P., Martens, K., de Olano, D., & Popp, M. (Eds.). (2010). *Das PISA-Echo*. Frankfurt am Main: Campus Verlag.
Knoepfel, P., Larrue, C., Varone, F., & Hill, M. (2007). *Public policy analysis*. Bristol: The Policy Press.
Kock, C. (2007). Is practical reasoning presumptive? *Informal Logic*, *27*(1), 91–108. Retrieved from http://hrgpapers.uwindsor.ca/ojs/leddy/index.php/informal_logic/article/viewArticle/466
Kreiner, S., & Christensen, K. B. (2014). Analyses of model fit and robustness. A new look at the PISA scaling model underlying ranking of countries according to reading literacy. *Psychometrika*, *79*(2), 210–31. http://doi.org/10.1007/s11336-013-9347-z
Krippendorff, K. (2004). *Content analysis: an introduction to its methodology* (2nd ed.). Thousand Oaks, CA: SAGE Publications, Inc.
Kvale, S. (2007). *Doing interviews*. London: Sage Publications Ltd.
Ladd, H. F. (2003). Comment on Caroline M. Hoxby: School choice and school competition: Evidence from the United States. *Swedish Economic Policy Review*, *10*, 67–76.
Ladwig, J. G. (2010). Beyond Academic Outcomes. *Review of Research in Education*, *34*(1), 113–141. http://doi.org/10.3102/0091732X09353062
Lafontaine, D. (2001a). *Au-delà des performances des jeunes de 15 ans, un système éducatif se profile... : Premiers résultats de PISA 2000*.
Lafontaine, D. (2001b). Le dispositif d'évaluation externe en Communauté française de Belgique. *Cahiers Du Service de Pédagogie Expérimentale*, *7-8*, 195–218.
Lafontaine, D. (2002). Le miracle "flamand" n'est pas prêt-à-porter…. *Cahiers Du Service de Pédagogie Expérimentale*, *9-10*, 223–226.
Lafontaine, D. (2007). D'un dispositif d'évaluations externes centré sur l'enseignant à un dispositif centré sur l'établissement et le système: le cas de la Communauté française de Belgique. *Journal of Educational Measurement and Applied Cognitive Sciences*, *1*(1), 1–11.
Lafontaine, D. (2014). À petits pas dans la bonne direction. *TRACeS de ChanGements*, *215*(March-April), 4–5.
Lafontaine, D., & Baye, A. (2012). PISA, instrument ou témoin du changement: évolution des performances en lecture et des politiques éducatives dans cinq systèmes européens.

Education Comparée, *7*.
Lafontaine, D., Baye, A., Burton, R., Demonty, I., Matoul, A., & Monseur, C. (2003). Les compétences des jeunes de 15 ans en Communauté française en lecture, en mathématiques et en sciences: Résultats de l'enquête PISA 2000. *Cahiers Du Service de Pédagogie Expérimentale*, *13-14*. Retrieved from http://www.oecd.org/belgium/33683502.pdf
Lafontaine, D., & Blondin, C. (2004). *Regards sur les résultats des élèves en Communauté française: apports des enquêtes de l'I.E.A., de PISA et des évaluations externes*. Bruxelles: De Boeck Supérieur.
Lafontaine, D., & Goffin, C. (2009). *Promouvoir l'orientation des filles vers les options scientifiques dès l'enseignement secondaire* (Final report). Liège: Service d'Analyse des Systèmes et des Pratiques d'Enseignement, Université de Liège.
Lafontaine, D., & Monseur, C. (2009). Gender gap in comparative studies of reading comprehension: To what extent do the test characteristics make a difference? *European Educational Research Journal*, *8*(1), 69–79. http://doi.org/10.2304/eerj.2009.8.1.69
Lapointe, A., Askew, J., & Mead, N. (1992). *Learning Science*. Princeton, NJ: Educational Testing Service.
LaRoche, S., Cartwright, F., & Statistics Canada. (2010). *Independent Review of the 2009 PISA Results for Ireland*. Dublin: Department of Education and Skills.
Lascoumes, P., & Le Galès, P. (2007). Introduction: Understanding Public Policy through Its Instruments-From the Nature of Instruments to the Sociology of Public Policy Instrumentation. *Governance*, *20*(1), 1–21. http://doi.org/10.1111/j.1468-0491.2007.00342.x
Lau, K.-C. (2009). A Critical Examination of PISA's Assessment on Scientific Literacy. *International Journal of Science and Mathematics Education*, *7*(6), 1061–1088. http://doi.org/10.1007/s10763-009-9154-2
Le Hebel, F., Montpied, P., & Tiberghien, A. (2014). Which Effective Competencies Do Students Use in PISA Assessment of Scientific Literacy? In C. Bruguière, A. Tiberghien, & P. Clément (Eds.), *Topics and Trends in Current Science Education* (Vol. 1, pp. 273–289). Dordrecht: Springer Science+Business Media B.V. http://doi.org/10.1007/978-94-007-7281-6_17
Le, L. T. (2009). Investigating Gender Differential Item Functioning Across Countries and Test Languages for PISA Science Items. *International Journal of Testing*, *9*(2), 122–133. http://doi.org/10.1080/15305050902880769
Leeuw, F. L. (2003). Reconstructing Program Theories: Methods Available and Problems to be Solved. *American Journal of Evaluation*, *24*(1), 5–20. http://doi.org/10.1016/S1098-2140(02)00271-0
Legewie, J., & DiPrete, T. A. (2012). School Context and the Gender Gap in Educational Achievement. *American Sociological Review*, *77*(3), 463–485. http://doi.org/10.1177/0003122412440802
Leithwood, K., & Jantzi, D. (2009). A Review of Empirical Evidence About School Size Effects: A Policy Perspective. *Review of Educational Research*, *79*(1), 464–490. http://doi.org/10.3102/0034654308326158
Lenkeit, J., & Caro, D. H. (2014). Performance status and change: measuring education system effectiveness with data from PISA 2000–2009. *Educational Research and Evaluation*, *20*(2), 146–174. http://doi.org/10.1080/13803611.2014.891462
Levin, B. (2005). *Reforming Education: From Origins to Outcomes*. Taylor & Francis e-Library.
Levin, B., Glaze, A., & Fullan, M. (2008). Results without Rancor or Ranking: Ontario's Success Story. *Phi Delta Kappan*, *90*(4), 273–280.
Lewis, M., & Archer, P. (2002). *Further Evaluation of Early Start: Progress Report*. Dublin: Educational Research Centre.
Lewis, M., Shortt, E., & Archer, P. (2011). Achievements of Early Start pupils in junior–infant classes. *The Irish Journal of Education*, *39*, 19–43.

Lietz, P. (2006a). A meta-analysis of gender differences in reading achievement at the secondary school level. *Studies in Educational Evaluation*, *32*(4), 317–344. http://doi.org/10.1016/j.stueduc.2006.10.002

Lietz, P. (2006b). Issues in the Change in Gender Differences in Reading Achievement in Cross-National Research Studies since 1992: A Meta-Analytic View. *International Education Journal*, *7*(2), 127–149.

Lingard, B. (2011). Policy as numbers: Ac/counting for educational research. *Australian Educational Researcher*, *38*(4), 355–382. http://doi.org/10.1007/s13384-011-0041-9

Link, S. (2012). *Developing key skills: What can we learn from various national approaches?* (EENEE Analytical Report No. 14). European Commission. Retrieved from http://www.eenee.de/portal/page/portal/EENEEContent/_IMPORT_TELECENTRUM/DOCS/EENEE_AR14.pdf

Loeb, S., & Strunk, K. (2007). Accountability and Local Control: Response to Incentives with and without Authority over Resource Generation and Allocation. *Education Finance and Policy*, *2*(1), 10–39. http://doi.org/10.1162/edfp.2007.2.1.10

Loerke, K. (2012). *Persistent and Contradictory Comparative Claims of Boys' and Girls' Reading Achievement: A Historical Interpretive Approach*. University of Alberta. Retrieved from http://hdl.handle.net/10402/era.29148

Lubienski, S. T. (2011). Mathematics Education and Reform in Ireland: An Outsider's Analysis of Project Maths. *Irish Mathematical Society Bulletin*, *67*, 27–55.

Lyons, M., Lynch, K., Close, S., Sheerin, E., & Boland, P. (2003). *Inside classrooms: The teaching and learning of mathematics in social context*. Dublin: Institute of Public Administration.

Lysaght, Z., & O'Leary, M. (2013). An instrument to audit teachers' use of assessment for learning. *Irish Educational Studies*, *32*(2), 217–232. http://doi.org/10.1080/03323315.2013.784636

Maaz, K., Trautwein, U., Lüdtke, O., & Baumert, J. (2008). Educational Transitions and Differential Learning Environments: How Explicit Between-School Tracking Contributes to Social Inequality in Educational Outcomes. *Child Development Perspectives*, *2*(2), 99–106. http://doi.org/10.1111/j.1750-8606.2008.00048.x

Mac Giolla Phádraig, B. (2010). Towards partnership: the development of parental involvement in Irish primary education since 1970. *The Irish Journal of Education*, *38*, 73–93.

Mahon, R. (2010). After Neo-Liberalism?: The OECD, the World Bank and the Child. *Global Social Policy*, *10*(2), 172–192. http://doi.org/10.1177/1468018110366615

Majone, G. (1996). *Temporal Consistency and Policy Credibility: Why Democracies Need Non-Majoritarian Institutions* (Working Paper No. 96/57). European University Institute. Retrieved from http://www.eui.eu/Documents/RSCAS/Publications/WorkingPapers/9657.pdf

Mangez, C., Maroy, C., Cattonar, B., Delvaux, B., & Mangez, E. (2009). *La construction des politiques de pilotage et d'évaluation en Communauté Française de Belgique: une approche cognitive*. Retrieved from http://www.knowandpol.eu/fileadmin/KaP/content/Scientific_reports/Orientation2/WP10_O2PA1_Rapport_court.pdf

Mangez, E. (2004). La production des programmes de cours par les agents intermédiaires: transfert de savoirs et relations de pouvoir. *Revue Française de Pédagogie*, *146*, 65–77.

Mangez, E., & Cattonar, B. (2009). The status of PISA in the relationship between civil society and the educational sector in French-speaking Belgium. *Sísifo. Educational Sciences Journal*, *10*, 15–26.

Mangez, E., & Hilgers, M. (2012). The field of knowledge and the policy field in education: PISA and the production of knowledge for policy. *European Educational Research Journal*, *11*(2), 189–205. http://doi.org/10.2304/eerj.2012.11.2.189

Maniquet, F. (2009). Inscriptions dans les écoles: quelques enjeux et quelques solutions.

Regards Économiques, *70*, 1–12.
Mann, A., & DiPrete, T. A. (2013). Trends in gender segregation in the choice of science and engineering majors. *Social Science Research*, *42*(6), 1519–1541. http://doi.org/10.1016/j.ssresearch.2013.07.002
Marcussen, M. (2004). Multilateral Surveillance and the OECD: Playing the Idea Game. In K. Armingeon & M. Beyeler (Eds.), *The OECD and European Welfare States* (pp. 13–31). Cheltenham: Edward Elgar. http://doi.org/10.4337/9781843769705.00008
Marcussen, M., & Trondal, J. (2011). The OECD civil servant: Caught between Scylla and Charybdis. *Review of International Political Economy*, *18*(5), 592 – 621. http://doi.org/10.1080/09692290.2011.603665
Marks, G. N. (2008). Accounting for the gender gaps in student performance in reading and mathematics: evidence from 31 countries. *Oxford Review of Education*, *34*(1), 89–109. http://doi.org/10.1080/03054980701565279
Martens, K., & Jakobi, A. P. (Eds.). (2010). *Mechanisms of OECD governance: International Incentives for National Policy-Making?* New York, NY: Oxford University Press.
Martens, K., Nagel, A.-K., Windzio, M., & Weymann, A. (Eds.). (2010). *Transformation of Education Policy*. Basingstoke: Palgrave Macmillan. http://doi.org/10.1057/9780230281295
Martens, K., & Niemann, D. (2010). *Governance by comparison: How ratings & rankings impact national policy-making in education* (TranState working papers No. 139). Retrieved from http://hdl.handle.net/10419/41595
Martens, K., & Niemann, D. (2013). When Do Numbers Count? The Differential Impact of the PISA Rating and Ranking on Education Policy in Germany and the US. *German Politics*, *22*(3), 314–332. http://doi.org/10.1080/09644008.2013.794455
Martens, K., Rusconi, A., & Leuze, K. (2007). *New arenas of education governance : the impact of international organizations and markets on educational policy making*. Basingstoke: Palgrave Macmillan.
Martin, M. (1997). Written Answers: Numeracy and Literacy Tests. *Dáil Éireann Debate*. 483(2), 20 November 1997. Retrieved from http://oireachtasdebates.oireachtas.ie/debates authoring/debateswebpack.nsf/takes/dail1997112000115?opendocument
Martin, M. (1998). Written Answers: Literacy Levels. *Dáil Éireann Debate*. 495(6), 22 October 1998.
Martin, M. (1999a). Written Answers: Science Education. *Dáil Éireann Debate*. 502(3), 23 March 1999. Retrieved from http://oireachtasdebates.oireachtas.ie/debates authoring/debateswebpack.nsf/takes/dail1999032300055?opendocument
Martin, M. (1999b). Written Answers: Science Education. *Dáil Éireann Debate*. 509(3), 14 October 1999. Retrieved from http://oireachtasdebates.oireachtas.ie/debates authoring/debateswebpack.nsf/takes/dail1999101400050?opendocument
Martin, M. O., Hickey, B. L., & Murchan, D. P. (1992). The Second International Assessment of Educational Progress: Mathematics And Science Findings In Ireland. *Irish Journal of Education*, *26*.
Martin, M. O., Mullis, I. V. S., Beaton, A. E., Gonzalez, E. J., Smith, T. A., & Kelly, D. L. (1997). *Science Achievement in the Primary School Years*. Chestnut Hill, MA: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.
Master, B. (2014). Staffing for Success: Linking Teacher Evaluation and School Personnel Management in Practice. *Educational Evaluation and Policy Analysis*, *36*(2), 207–227. http://doi.org/10.3102/0162373713506552
Mathematics Expert Group. (2010). *PISA 2012 Mathematics Framework* (Draft to OECD). Retrieved from https://www.oecd.org/pisa/pisaproducts/46961598.pdf
Mayne, J. (2012). Contribution analysis: Coming of age? *Evaluation*, *18*(3), 270–280. http://doi.org/10.1177/1356389012451663
Mazumdar, K. (2000). Causal flow between human well-being and per capita real gross

domestic product. *Social Indicators Research*, *50*, 297–313. http://doi.org/10.1023/A:1007061703600
Mazzeo, J., & von Davier, M. (2008). *Review of the Programme for International Student Assessment (PISA) test design: Recommendations for fostering stability in assessment results.* (No. EDU/PISA/GB(2008)28). Princeton, NJ.
McAvinue, L. (2014). Trends over Time: Achievement & Retention Levels. In *Learning from DEIS* (pp. 76–95). Dublin. Retrieved from http://www.education.ie/en/Schools-Colleges/Services/DEIS-Delivering-Equality-of-Opportunity-%E2%80%8Ein-Schools-/DEIS-Supporting-Information/Learning-from-DEIS-Seminar.html
McCarthy, J., & Burns, D. (2005). Mathematics: the Cinderella subject of learning support. *InTouch*, (69), 28–29.
McElduff, A., & Sheehan, B. (2001). *The Education Act 1998 and the Education Welfare Act 2000: Question and Answer Guide*. Dublin: Irish National Teachers' Organisation.
McGaw, B. (2008). The Role of the OECD in International Comparative Studies of Achievement. *Assessment in Education: Principles, Policy & Practice*, *15*(3), 223–243.
McGill, R., Tukey, J. W., & Larsen, W. A. (1978). Variations of Box Plots. *The American Statistician*, *32*(1), 12–16.
McMorris, D. (2008). Second Level Teaching Standards: Discussion with Department of Education and Science. *Joint Committee on Education and Science Debate*. 6 November 2008. Retrieved from http://oireachtasdebates.oireachtas.ie/Debates Authoring/DebatesWebPack.nsf/committeetakes/EDJ2008110600003?opendocument
McNamara, G. (2006). Workable Compromise or Pointless Exercise?: School-based Evaluation in the Irish Context. *Educational Management Administration & Leadership*, *34*(4), 564–582. http://doi.org/10.1177/1741143206068218
McNamara, G., & O'Hara, J. (2005). Internal review and self-evaluation — the chosen route to school improvement in ireland? *Studies in Educational Evaluation*, *31*(4), 267–282. http://doi.org/10.1016/j.stueduc.2005.11.003
McNamara, G., & O'Hara, J. (2012). From Looking at Our Schools (LAOS) to Whole School Evaluation - Management, Leadership and Learning (WSE-MLL): The Evolution of Inspection in Irish Schools over the Past Decade. *Educational Assessment, Evaluation and Accountability*, *4*(2), 79–97.
McNamara, G., O'Hara, J., Lisi, P. L., & Davidsdottir, S. (2011). Operationalising self-evaluation in schools: experiences from Ireland and Iceland. *Irish Educational Studies*, *30*(1), 63–82. http://doi.org/10.1080/03323315.2011.535977
Meghir, C., & Rivkin, S. G. (2010). *Econometric Methods for Research in Education*. Retrieved from http://www.nber.org/papers/w16003
Mehana, M., & Reynolds, A. J. (2004). School mobility and achievement: A meta-analysis. *Children and Youth Services Review*, *26*, 93–119. http://doi.org/10.1016/j.childyouth.2003.11.004
Merrifield, J. (2013). Can We Reasonably Assess "Productivity of Market-Based Educational Reforms"?: Comment on Anna Egalite's "Measuring Competitive Effects From School Voucher Programs: A Systematic Review." *Journal of School Choice*, *7*(4), 465–467. http://doi.org/10.1080/15582159.2013.837763
Merriman, B., Shiel, G., Cosgrove, J., & Perkins, R. (2014). *Project Maths and PISA 2012*. Dublin: Educational Research Centre.
Meseguer, C., & Gilardi, F. (2009). What is new in the study of policy diffusion? *Review of International Political Economy*, *16*(3), 527–543. http://doi.org/10.1080/09692290802409236
Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York: American Council of Education/Macmillian.
Messick, S. (1998). Test Validity: A Matter of Consequence. *Social Indicators Research*, *45*, 35–44. http://doi.org/10.2307/27522333
Meyer, M. (2010). The Rise of the Knowledge Broker. *Science Communication*, *32*(1), 118–

127. http://doi.org/10.1177/1075547009359797
Michaelides, M. P., & Haertel, E. H. (2014). Selection of Common Items as an Unrecognized Source of Variability in Test Equating: A Bootstrap Approximation Assuming Random Sampling of Common Items. *Applied Measurement in Education*, *27*(1), 46–57. http://doi.org/10.1080/08957347.2013.853069
Ministère de l'Education. (1991). *Les systèmes éducatifs en Belgique: similitudes et divergences*. Bruxelles: Ministère de l'Education, de la Recherche et de la Formation.
Ministère de la Communauté française. (2008). *Le Développement de l'éducation: rapport de la Communauté française de Belgique*. Bruxelles: Ministère de la Communauté française. Retrieved from http://www.ibe.unesco.org/National_Reports/ICE_2008/belgium-cfwb_NR08_fr.pdf
Ministère de la Communauté française. (2009a). Premier degré de l'enseignement secondaire. Retrieved from http://www.enseignement.be/download.php?do_id=4180&do_check=
Ministère de la Communauté française. (2009b). *Premier degré de l'enseignement secondaire: expos des motifs*. Retrieved from http://www.enseignement.be/download.php?do_id=4180&do_check=
Ministère de la Communauté française. (2014). *Core Skills*. Retrieved from http://www.enseignement.be/index.php?page=24737&navi=295
Ministère de la Communauté française. (2015). Évaluation certificative - Certificat d'Études de Base (CEB). Retrieved September 22, 2015, from http://www.enseignement.be/index.php?page=25527
Mitchell, L. (2011). Enquiring teachers and democratic politics: transformations in New Zealand's early childhood education landscape. *Early Years*, *31*(3), 217–228. http://doi.org/10.1080/09575146.2011.588787
MOH Portal Team. (2014, May 13). The MOH, in Collaboration with the Ministry of Education, Evaluates the Role of the Health Affairs Directorates in Educating on MERS CoronaVirus. *Ministry News*. Retrieved from http://www.moh.gov.sa/en/Ministry/MediaCenter/News/Pages/News-2014-05-13-002.aspx
Moldovan Association of ICT Companies. (2014). Collaboration between the Ministry of Information Technology and Communications, the Ministry of Education and IBM for the educational system development. *News*. Retrieved from http://www.ict.md/2014-07-03
Moldoveanu, M. C., & Baum, J. A. C. (2005). Contemporary Debates in Organisational Epistemology. In J. A. C. Baum (Ed.), *Companion to Organizations* (pp. 733–751). Malden, MA: Blackwell Publishing Ltd.
Monseur, C. (2000). Résultats d'une enquête internationale en mathématiques. *Cahiers Du Service de Pédagogie Expérimentale*, *3-4*, 13–20.
Monseur, C., Baye, A., Lafontaine, D., & Quittre, V. (2011). PISA test format assessment and the local independence assumption. In *Ieri Monographs Series. Issues and Methodologies in Large-Scale assessments.* (Vol. 4, pp. 131–158).
Monseur, C., & Berezner, A. (2007). The computation of equating errors in international surveys in education. *Journal of Applied Measurement*, *8*(3), 323–35.
Monseur, C., & Lafontaine, D. (2009). L'organisation des systèmes éducatifs: quel impact sur l'efficacité et l'équité? In V. Dupriez & X. Dumay (Eds.), *L'efficacité en éducation, promesses et zones d'ombre* (pp. 141–163). Bruxelles: De Boeck.
Morgan, C. (2011). Constructing the OECD Programme for International Student Assessment. In M. A. Pereyra, H.-G. Kotthoff, & R. Cowen (Eds.), *PISA under examination: Changing knowledge, changing tests, and changing schools* (pp. 47–59). Rotterdam: SensePublishers.
Morgan, C., & Shahjahan, R. A. (2014). The legitimation of OECD's global educational governance: examining PISA and AHELO test production. *Comparative Education*, *50*(2), 192–205. http://doi.org/10.1080/03050068.2013.834559

Morgan, G. (2006). *Images of Organization* (Updated). Thousand Oaks, CA: SAGE Publications, Inc.
Morgan, S. L., Gelbgiser, D., & Weeden, K. A. (2013). Feeding the pipeline: Gender, occupational plans, and college major selection. *Social Science Research*, *42*(4), 989–1005. http://doi.org/10.1016/j.ssresearch.2013.03.008
Muijs, D., Kyriakides, L., van der Werf, G., Creemers, B. P. M., Timperley, H., & Earl, L. (2014). State of the art – teacher effectiveness and professional learning. *School Effectiveness and School Improvement*, *25*(2), 231–256. http://doi.org/10.1080/09243453.2014.885451
Mujtaba, T., & Reiss, M. J. (2013). What Sort of Girl Wants to Study Physics After the Age of 16? Findings from a Large-scale UK Survey. *International Journal of Science Education*, *35*(17), 2979–2998. http://doi.org/10.1080/09500693.2012.681076
Mullis, I. V. S., & Martin, M. O. (2006). *TIMSS in Perspective: Lessons Learned from IEA's Four Decades of International Mathematics Assessments* (Draft). *2nd IEA International Research Conference (IRC-2006)*. Washington, DC. Retrieved from http://www.iea.nl/irc-2006.html
Mullis, I. V. S., Martin, M. O., Beaton, A. E., Gonzalez, E. J., Kelly, D. L., & Smith, T. A. (1997). *Mathematics Achievement in the Primary School Years*. Chestnut Hill, MA: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.
Mullis, I. V. S., Martin, M. O., Beaton, A. E., Gonzalez, E. J., Kelly, D. L., Smith, T. A., & Hall, C. (1998). *Mathematics and Science Achievement in the Final Year of Secondary School*. Chestnut Hill, MA: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.
Mullis, I. V. S., Martin, M. O., Foy, P., & Arora, A. (2012). *TIMSS 2011 International Results in Mathematics*. Chestnut Hill, MA, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.
Mullis, I. V. S., Martin, M. O., Foy, P., Olson, J. F., Preuschoff, C., Erberber, E., … Galia, J. (2008). *TIMSS 2007 International Mathematics Report*. Chestnut Hill, MA.
Mullis, I. V. S., Martin, M. O., Minnich, C. A., Stanco, G. M., Arora, A., Centurino, V. A. S., & Castle, C. E. (Eds.). (2012). *TIMSS 2011 Encyclopedia: Education Policy and Curriculum in Mathematics and Science* (Vol. 1: A-K). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
Murphy, C. (2013). Science items: Context and curriculum. In *National Schools, international contexts: Beyond the PIRLS and TIMSS test results* (pp. 177–200). Dublin: Educational Research Centre.
Murphy, C., & Smith, G. (2012). The impact of a curriculum course on pre-service primary teachers' science content knowledge and attitudes towards teaching science. *Irish Educational Studies*, *31*(1), 77–95. http://doi.org/10.1080/03323315.2011.634061
Murphy, D. (2014). Issues with PISA's Use of its Data in the Context of International Education Policy Convergence. *Policy Futures in Education*, *12*(7), 893–916. http://doi.org/10.2304/pfie.2014.12.7.893
Murtagh, E. (2008). Concluding Remarks. In *PISA National Symposium*. Educational Research Centre. Retrieved from http://www.erc.ie/?p=139
Musset, P. (2012). *School Choice and Equity: Current Policies in OECD Countries and a Literature Review* (No. 66). *OECD Education Working Papers*. OECD Publishing.
Nagel, A.-K., Bieber, T., Jakobi, A. P., Knodel, P., Niemann, D., & Teltemann, J. (2010). Measuring Transformation of Education Policy—a Mixed-Method Approach. In K. Martens, A.-K. Nagel, M. Windzio, & A. Weymann (Eds.), *Transformation of Education Policy* (pp. 28–50). Basingstoke: Palgrave Macmillan.
National Commission on Excellence in Education. (1983). *A Nation at Risk: the imperative for educational reform*. Washington, DC: U.S. Government Printing Office.
National Council for Special Education. (2013). Members of the National Council for Special Education: Eamon Stack. Retrieved September 3, 2015, from http://ncse.ie/members-of-the-national-council-for-special-education

National Educational Welfare Board. (2013). *Home School Community Liaison Scheme*. Department of Education and Skills. Retrieved from http://www.newb.ie/publications/hscldocumentation.asp
National Research Council. (2012). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. Washington, DC: National Academy Press.
NCCA. (1993). *Curriculum and Assessment Policy: Towards the new century*. Dublin: National Council for Curriculum and Assessment.
NCCA. (1999). *Primary School Curriculum: Introduction*. Dublin: The Stationery Office.
NCCA. (2003). Notes on the revised syllabus. Retrieved from http://www.pdst.ie/sites/default/files/jc_science_sy_rev_notes.pdf
NCCA. (2005). *Review of Mathematics in Post-Primary Education* (Discussion Paper). Dublin: NCCA.
NCCA. (2006). *Review of Mathematics in Post-Primary Education* (Report on the Consultation). Dublin: NCCA. Retrieved from http://ncca.ie/en/Curriculum_and_Assessment/Post-Primary_Education/Project_Maths/Information/Reports/
NCCA. (2007). *Assessment in the Primary School Curriculum: Guidelines for Schools*. Dublin: NCCA.
NCCA. (2008). *Junior Certificate: Science Syllabus (Ordinary Level and Higher Level)*. Dublin: The Stationery Office.
NCCA. (2009). *Code of Practice for the Governance of the National Council for Curriculum and Assessment*. National Council for Curriculum and Assessment. Retrieved from http://ncca.ie/en/About_Us/Governance/
NCCA. (2011). *Project Maths: Reviewing the project in the initial group of 24 schools – report on school visits*. National Council for Curriculum and Assessment. Retrieved from http://ncca.ie/en/Curriculum_and_Assessment/Post-Primary_Education/Project_Maths/Information/Reports/
NCCA. (2012). *Project Maths: Responding to current debate*. NCCA. Retrieved from http://ncca.ie/en/Curriculum_and_Assessment/Post-Primary_Education/Project_Maths/Information/Reports/
NCCA. (2014). *Maths in Practice: Report and recommendations*. National Council for Curriculum and Assessment. Retrieved from http://ncca.ie/en/Curriculum_and_Assessment/Post-Primary_Education/Project_Maths/Information/Reports/
NCCA. (2015). Project Maths. Retrieved September 2, 2015, from http://www.ncca.ie/en/Curriculum_and_Assessment/Post-Primary_Education/Project_Maths/
Negrea, G. O., & Duse, D.-M. (2014). Investigating autonomy, accountability and educational management priorities in romanian high schools; preliminary results. In *Management challenges for sustainable development* (pp. 35–50). Bucharest.
Neumann, E., Kiss, A., & Fejes, I. (2012). The Hard Work of Interpretation: The National Politics of PISA Reception in Hungary and Romania. *European Educational Research Journal*, *11*(2), 227–242.
Neumann, K., Fischer, H. E., & Kauertz, A. (2010). From PISA to educational standards: The impact of large-scale assessments on science education in Germany. *International Journal of Science and Mathematics Education*, *8*, 545–563. http://doi.org/10.1007/s10763-010-9206-7
Newton, P. E., & Shaw, S. D. (2014). *Validity in educational and psychological assessment*. London: SAGE Publications Ltd.
Nilsson, M., Zamparutti, T., Petersen, J. E., Nykvist, B., Rudberg, P., & Mcguinn, J. (2012). Understanding Policy Coherence: Analytical Framework and Examples of Sector-Environment Policy Interactions in the EU. *Environmental Policy and Governance*, *22*, 395–423. http://doi.org/10.1002/eet.1589
Niukko, S. (2006). OECD in the eyes of Finnish education policymakers. In J. Kallo & R.

Rinne (Eds.), *Supranational Regimes and National Education Policies: Encountering Challenges* (pp. 299–333). Turku: Finnish Educational Research Association, Kasvatusalan tutkimuksia.
Nóvoa, A., & Yariv-Mashal, T. (2003). Comparative Research in Education: a mode of governance or a historical journey? *Comparative Education, 39*(4), 423–438. http://doi.org/10.1080/0305006032000162002
Nuffield Foundation. (2015). Welcome to Twenty First Century Science. Retrieved from http://www.nuffieldfoundation.org/twenty-first-century-science
O'Connor, M. (2014). Investment in edification: reflections on Irish education policy since independence. *Irish Educational Studies*, *33*(2), 193–212. http://doi.org/10.1080/03323315.2014.920609
O'Donoghue, J. (2002). Numeracy and Mathematics. *Irish Mathematical Society Bulletin, 48*, 47–55.
O'Donoghue, T., & Harford, J. (2011). A Comparative History of Church-State Relationsin Irish Education. *Comparative Education Review*, *55*(3), 315–341. http://doi.org/10.1086/659871
O'Keeffe, L., & O'Donoghue, J. (2011). *A Review of School Textbooks for Project Maths*. Limerick: National Centre for Excellence in Mathematics and Science Teaching and Learning.
O'Leary, M. (2001). The Effects of Age-based and Grade-based Sampling on the Relative Standing of Countries in International Comparative Studies of Student Achievement. *British Educational Research Journal*, *27*(2), 187–200. http://doi.org/10.1080/01411920120037135
O'Reilly, M. (2006). Irish Matematical Society: President's report 2006. *Irish Mathematical Society Bulletin, 58*, 9–12.
O'Sullivan, D. (2005). *Cultural Politics and Irish Education since the 1950s: Policy paradigms and power*. Dublin: Institute of Public Administration.
OCDE. (1999). *Mesurer les connaissances et compétences des élèves: Un nouveau cadre d'évaluation*. Paris: Éditions OCDE. http://doi.org/10.1787/9789264273122-fr
Odrovic, M. (2003). La pédagogie des compétences. *Informations Pédagogiques*, *54*, 3–10.
OECD. (1969). *Reviews of National Policies for Education: Ireland*. Paris: OECD Publishing.
OECD. (1978). *Declaration on Future Educational Policies in the Changing Social and Economic Context* (ED/MIN(78)4/FINAL). Retrieved from http://acts.oecd.org/Instruments/ShowInstrumentView.aspx?InstrumentID=144&InstrumentPID=140&Lang=en
OECD. (1991). *Reviews of National Policies for Education: Ireland*. Paris: OECD Publishing.
OECD. (1993). *Reviews of National Policies for Education: Belgium*. Paris: OECD Publishing.
OECD. (1996). *The Knowledge-Based Economy*. Paris: OECD Publishing. Retrieved from http://www.oecd.org/dataoecd/51/8/1913021.pdf
OECD. (1997a). First meeting of the Board of Participating Countries. BPC(97.1).99bis.
OECD. (1997b). OECD Programme for Producing Student Achievement Indicators on a Regular Basis: Second meeting of the Board of Participating Countries. BPC(97.2).99.
OECD. (1997c). *Parents as Partners in Schooling*. Paris: OECD Publishing.
OECD. (1997d). Resolution of the Council on a decentralised Programme for Producing Indicators on Student Achievement on a Regular Basis. C(97)176/FINAL.
OECD. (1998a). OECD Programme for International Student Assessment (PISA): Fourth meeting of the Board of Participating Countries (BPC). DEELSA/PISA/BPC(98)16.
OECD. (1998b). OECD Programme for International Student Assessment: Fifth meeting of the Board of Participating Countries. DEELSA/PISA/BPC(98)32.
OECD. (1998c). Programme for Producing Student Achievement Indicators on a Regular Basis: Third meeting of the Board of Participating Countries (BPC).

DEELSA/PISA/BPC(98)8.
OECD. (1999a). *Measuring student knowledge and skills: a new framework for assessment*. Paris: OECD Publishing.
OECD. (1999b). OECD programme for international student assessment (PISA): Seventh meeting of the Board of Participating Countries. DEELSA/PISA/BPC/M(99)1.
OECD. (1999c). OECD Programme for International Student Assessment: Sixth meeting of the Board of Participating Countries. DEELSA/PISA/BPC/M(99)1.
OECD. (2001a). *Knowledge and Skills for Life: first results from the OECD Programme for International Student Assessment (PISA) 2000*. Paris: OECD Publishing.
OECD. (2001b). OECD Programme for International Student Assessment: Eleventh meeting of the Board of Participating Countries. DEELSA/PISA/BPC/M(2001)2.
OECD. (2001c). OECD Programme for International Student Assessment: Tenth meeting of the Board of Participating Countries. DEELSA/PISA/BPC/M(2001)1.
OECD. (2002a). OECD programme for international student assessment (PISA): Fourteenth meeting of the Board of Participating Countries. DEELSA/PISA/BPC/M(2002)2/REV1.
OECD. (2002b). OECD Programme for International Student Assessment: Thirteenth meeting of the Board of Participating Countries. DEELSA/PISA/BPC/M(2002)1/REV1.
OECD. (2002c). OECD Programme for International Student Assessment: Twelfth meeting of the Board of Participating Countries. DEELSA/PISA/BPC/M(2001)3/REV1.
OECD. (2003a). Foreword. In *Student engagement at school: A sense of belonging and participation* (pp. 3–4). Paris: OECD Publishing.
OECD. (2003b). OECD programme for international student assessment (PISA): Fifteenth meeting of the Board of Participating Countries. EDU/PISA/BPC/M(2003)1.
OECD. (2003c). OECD Programme for International Student Assessment (PISA): Sixteenth meeting of the Board of Participating Countries. EDU/PISA/BPC/M(2003)2.
OECD. (2004a). *First results from PISA 2003: Executive Summary*. Paris: OECD Publishing.
OECD. (2004b). *Learning for Tomorrow's World: First Results from PISA 2003*. Paris: OECD Publishing.
OECD. (2005a). Longer term strategy of the development of PISA: 20th meeting of the PISA Governing Board. EDU/PISA/GB(2005)21.
OECD. (2005b). *PISA 2003 Technical Report*. Paris: OECD Publishing.
OECD. (2005c). *Teachers Matter: Attracting, Developing and Retaining Effective Teachers*. Paris: OECD Publishing. http://doi.org/10.1787/9789264018044-en
OECD. (2006). *Assessing Scientific, Reading and Mathematical Literacy: A Framework for PISA 2006*. Paris: OECD Publishing.
OECD. (2007a). *PISA 2006* (Vol. 1). OECD Publishing. http://doi.org/10.1787/9789264040014-en
OECD. (2007b). *PISA 2006: Science Competencies for Tomorrow's World* (Vol. 1:Analysis). OECD Publishing. http://doi.org/10.1787/9789264040151-en
OECD. (2007c). *PISA 2006: Science Competencies for Tomorrow's World - Executive Summary*. Retrieved from https://www.oecd.org/pisa/pisaproducts/39725224.pdf
OECD. (2007d). Strategy for the Evaluation of the Policy Impact of PISA: 24th meeting of the PISA Governing Board. EDU/PISA/GB(2007)35.
OECD. (2008a). *OECD Public Management Reviews: Ireland 2008*. Paris: OECD Publishing.
OECD. (2008b). *The OECD: Organisation for Economic Co-operation and Development*. Paris: OECD Publishing. Retrieved from https://www.oecd.org/newsroom/34011915.pdf
OECD. (2009a). *Creating effective teaching and learning environments: First results from TALIS*. Paris: OECD Publishing.
OECD. (2009b). OECD Programme for International Student Assessment: 27th meeting of

the PISA Governing Board. EDU/PISA/GB/M(2009)1/REV1.
OECD. (2009c). *PISA 2006 Technical Report*. Paris: OECD Publishing.
OECD. (2009d). *PISA Data Analysis Manual* (SPSS, 2nd). Paris: OECD Publishing.
OECD. (2009e). *Top of the Class: High Performers in Science in PISA 2006*. Paris: OECD Publishing. Retrieved from http://www.oecd-ilibrary.org/education/top-of-the-class_9789264060777-en
OECD. (2010a). OECD Programme for International Student Assessment: 30th meeting of the PISA Governing Board (Draft summary record). EDU/PISA/GB/M(2010)2/REV1.
OECD. (2010b). *PISA 2009 Results: Executive Summary*.
OECD. (2010c). *PISA 2009 Results: Learning to Learn* (Vol. III). Paris: OECD Publishing.
OECD. (2010d). *PISA 2009 Results: Learning Trends* (Vol. V). Paris: OECD Publishing.
OECD. (2010e). *PISA 2009 Results: Overcoming Social Background* (Vol. II). Paris: OECD Publishing.
OECD. (2010f). *PISA 2009 Results: What Makes a School Successful?* (Vol. IV). Paris: OECD Publishing.
OECD. (2010g). *PISA 2009 Results: What Students Know and Can Do* (Vol. I). Paris: OECD Publishing.
OECD. (2011a). *Better Policies for Better Lives: The OECD at 50 and beyond*. Paris: OECD Publishing.
OECD. (2011b). Does participation in pre-primary education translate into better learning outcomes at school? *PISA in Focus*, *1*. http://doi.org/10.1787/5k9h362tpvxp-en
OECD. (2011c). Improving Performance: Leading from the Bottom. *PISA in Focus*, (2).
OECD. (2011d). *Learning for jobs: Pointers for policy development* (OECD reviews of vocational education and training). Paris: OECD Publishing. Retrieved from https://www.oecd.org/education/innovation-education/learningforjobs.htm
OECD. (2011e). PISA Longer Term Strategy Papers: Meeting of the PISA Strategic Development Group. EDU/PISA/GB/SDG(2011)5.
OECD. (2011f). School autonomy and accountability: Are they related to student performance? *Pisa in Focus*, (9).
OECD. (2011g). *Teachers Matter: Attracting, Developing and Retaining Effective Teachers* (Pointers for policy development). Retrieved from https://www.oecd.org/edu/school/48627229.pdf
OECD. (2011h). *The Impact of the 1999 Education Reform in Poland* (OECD Education Working Papers No. 49). Paris: OECD Publishing.
OECD. (2011i). When students repeat grades or are transferred out of school: What does it mean for education systems? *PISA in Focus*, (6).
OECD. (2012a). Are school vouchers associated with equity in education? *PISA in Focus*, (20).
OECD. (2012b). Does performance-based pay improve teaching? *PISA in Focus*, *16*.
OECD. (2012c). *Equity and Quality in Education: Supporting Disadvantaged Students and Schools*. Paris: OECD Publishing. Retrieved from http://www.oecd-ilibrary.org/education/equity-and-quality-in-education_9789264130852-en
OECD. (2012d). *Indicators of Education Systems* (Online brochure). OECD Publishing. Retrieved from https://www.oecd.org/edu/skills-beyond-school/49338320.pdf
OECD. (2012e). *PISA 2009 Technical Report*. OECD Publishing. http://doi.org/10.1787/9789264167872-en
OECD. (2012f, February). Does Money Buy Strong Performance in PISA? *PISA in Focus*, *13*. http://doi.org/10.1787/5k9fhmfzc4xx-en
OECD. (2013a). Beyond PISA 2015: A Longer-Term Strategy of PISA - 36th meeting of the PISA Governing Board. EDU/PISA/GB(2013)14.
OECD. (2013b). *Education Policy Outlook: Australia*. Paris: OECD Publishing.
OECD. (2013c). *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*. Paris: OECD Publishing. http://doi.org/10.1787/9789264190511-en

OECD. (2013d). *PISA 2012 Results in Focus*. Retrieved from http://www.oecd.org/pisa/keyfindings/pisa-2012-results.htm
OECD. (2013e). *PISA 2012 Results: Excellence through Equity* (Vol. II). Paris: OECD Publishing. http://doi.org/10.1787/9789264201132-en
OECD. (2013f). *PISA 2012 Results: Ready to Learn* (Vol. III). Paris: OECD Publishing. http://doi.org/10.1787/9789264201170-en
OECD. (2013g). *PISA 2012 Results: What Makes Schools Successful?* (Vol. IV). Paris: OECD Publishing. Retrieved from http://www.oecd-ilibrary.org/education/pisa-2012-results-what-makes-a-school-successful-volume-iv_9789264201156-en
OECD. (2013h). Who are the strong performers and successful reformers in education? *PISA in Focus*, (34).
OECD. (2014a). Are disadvantaged students more likely to repeat grades? *PISA in Focus*, (43).
OECD. (2014b). Are grouping and selecting students for different schools related to students' motivation to learn? *PISA in Focus*, (39).
OECD. (2014c). Does pre-primary education reach those who need it most? *PISA in Focus*, (6). Retrieved from http://www.oecd.org/pisa/pisaproducts/pisainfocus.htm
OECD. (2014d). *Education at a Glance 2014: OECD Indicators*. Paris: OECD Publishing. Retrieved from http://www.oecd-ilibrary.org/education/education-at-a-glance-2014_eag-2014-en
OECD. (2014e). *New Insights from TALIS 2013: Teaching and Learning in Primary and Upper Secondary Education*. Paris: OECD Publishing. http://doi.org/10.1787/9789264226319-en
OECD. (2014f). *PISA 2012 Results: Creative Problem Solving* (Vol. V). Paris: OECD Publishing.
OECD. (2014g). *PISA 2012 Results: What Students Know and Can Do* (Revised, Vol. 1). OECD Publishing. http://doi.org/10.1787/9789264201118-en
OECD. (2014h). *PISA 2012: Technical Report*. Paris: OECD Publishing.
OECD. (2014i). *TALIS 2013 Results: An International Perspective on Teaching and Learning*. Paris: OECD Publishing. http://doi.org/10.1787/9789264196261-en
OECD. (2014j). The OECD's contribution on education to the post-2015 framework: PISA for development. Retrieved from https://www.oecd.org/pisa/aboutpisa/POST-2015 paper education_20 Feb.pdf
OECD. (2015a). Budget. Retrieved November 3, 2015, from https://www.oecd.org/about/budget/
OECD. (2015b). Decisions, Recommendations and other Instruments of the Organisation for Economic Co-Operation and Development. Retrieved December 5, 2015, from http://webnet.oecd.org/OECDACTS/Instruments/ListBySubjectView.aspx
OECD. (2015c). *Education Policy Outlook 2015: Making reforms happen*. Paris: OECD Publishing.
OECD. (2015d). Education Policy Outlook Reforms Finder. Retrieved April 16, 2016, from https://www.oecd.org/edu/reformsfinder.htm
OECD. (2015e). OECD Legal Instruments. Retrieved from https://www.oecd.org/legal/legal-instruments.htm
OECD. (2015f). *OECD work on Education & Skills*. Paris: OECD Publishing. Retrieved from https://www.oecd.org/edu/Directorate-for-education-and-skills-brochure.pdf
OECD. (2015g). On-Line Guide to OECD Intergovernmental Activity. Retrieved November 3, 2015, from http://webnet.oecd.org/oecdgroups/
OECD. (2015h). PISA FAQ. Retrieved from http://www.oecd.org/pisa/aboutpisa/pisafaq.htm
OECD. (2015i). *The ABC of Gender Equality in Education: Aptitude, Behaviour, Confidence*. Paris: OECD Publishing. http://doi.org/10.1787/9789264229945-en
OECD. (2015j). Who does what. Retrieved May 28, 2015, from https://www.oecd.org/about/whodoeswhat/

OECD. (2016a). Education & Skills Today. Retrieved April 16, 2016, from http://oecdeducationtoday.blogspot.co.uk/
OECD. (2016b). Education GPS: The world of education at your fingertips. Retrieved April 16, 2016, from http://gpseducation.oecd.org/
OECD. (2016c). OECD Observer. Retrieved April 16, 2016, from http://www.oecdobserver.org/
OECD Directorate for Education. (2004). *Early Childhood Education and Care Policy: Country note for Ireland*. Paris: OECD Publishing.
OECD, & Joint Research Centre of the European Commission. (2008). *Handbook on Constructing Composite Indicators: Methodology and user guide*. Paris: OECD Publishing.
OECD, & UNESCO-UIS. (2003a). *Literacy skills for the world of tomorrow: further results from PISA 2000*. Paris: UNESCO Publishing/OECD Publishing.
OECD, & UNESCO-UIS. (2003b). *Literacy skills for the world of tomorrow: Further results from PISA 2000* (Executive summary).
Ogura, Y., & Goto, K. (2013). *The Process of Becoming: Not Valuing Science Study during Secondary School in Japan* (Paper presented at the New Perspectives in Science Education conference,14-15 March). Florence. Retrieved from http://conference.pixel-online.net/npse2013/common/download/Paper_pdf/275-SEC04-FP-Ogura-NPSE2013.pdf
Oireachtas. Education act, 1998, Pub. L. No. 51 (1998). Ireland.
Oldham, E. (2001). The culture of mathematics education in the republic of Ireland: Keeping the faith? *Irish Educational Studies*, *20*, 266–277. http://doi.org/10.1080/0332331010200121
Oliveri, M. E., & von Davier, M. (2011). Investigation of model fit and score scale comparability in international assessments. *Psychological Test and Assessment Modeling*, *53*(3), 315–333. Retrieved from http://www.doaj.org/doaj?func=fulltext&aId=826247
Owen, E., Hodgkinson, G. D., & Tuijnman, A. (1995). Towards a Strategic Approach for Developing International Indicators of Student Achievement. In G. Phillips & J. H. Moskowitz (Eds.), *Measuring What Students Learn*. Paris: OECD Publishing.
Papadopoulos, G. S. (1995). *Education 1960-1990: The Oecd Perspective*. Paris: OECD Publishing.
Parlement de la Communauté Française. (1997). Séances du mercredi 21 mai 1997 (matin et après-midi). Compte rendu intégral (1996-1997) n° 17.
Patton, M. Q. (2002). *Qualitative Research & Evaluation Methods*. Thousand Oaks, CA: SAGE Publications, Inc.
Patton, M. Q. (2008). *Utilization-Focused Evaluation* (4th ed.). Thousand Oaks, CA: SAGE Publications. Retrieved from http://www.amazon.com/dp/141295861X
Pawson, R., & Tilley, N. (1997). *Realistic Evaluation*. Thousand Oaks, CA: SAGE Publications.
Pedró, F. (2012). Deconstruyendo los puentes de Pisa: del análisis de resultados a la prescripción política. *Revista Española de Educación Comparada*, *19*(2012), 139–172. Retrieved from http://www.researchgate.net/publication/234139385_Deconstruyendo_los_puentes_de_Pisa_del_anlisis_de_resultados_a_la_prescripcin_poltica
Perez-Felkner, L., McDonald, S.-K., Schneider, B., & Grogan, E. (2012). Female and male adolescents' subjective orientations to mathematics and the influence of those orientations on postsecondary majors. *Developmental Psychology*, *48*(6), 1658–1673. http://doi.org/10.1037/a0027020
Perkins, R., Cosgrove, J., Moran, G., & Shiel, G. (2012). *PISA 2009: Results for Ireland and Changes Since 2000*. Dublin: Educational Research Centre.
Perkins, R., Moran, G., Cosgrove, J., & Shiel, G. (2010). *PISA 2009: The Performance and Progress of 15-year-olds in Ireland* (Summary Report). Dublin: Educational Research

Centre.
Perkins, R., Shiel, G., Merriman, B., Cosgrove, J., & Moran, G. (2013). *Learning for Life: The Achievements of 15-year-olds in Ireland on Mathematics, Reading Literacy and Science in PISA 2012*. Dublin: Educational Research Centre.
Permani, R. (2009). The role of education in economic growth in East Asia: A survey. *Asian-Pacific Economic Literature*, *23*, 1–20. http://doi.org/10.1111/j.1467-8411.2009.01220.x
Pfeffer, F. T. (2008). Persistent inequality in educational attainment and its institutional context. *European Sociological Review*, *24*(5), 543–565. http://doi.org/10.1093/esr/jcn026
Phillips, D., & Ochs, K. (2004). Researching policy borrowing: Some methodological challenges in comparative education. *British Educational Research Journal*, *30*(6), 773–784. http://doi.org/10.1080/0141192042000279495
Pianta, R. C., Barnett, W. S., Burchinal, M., & Thornburg, K. R. (2009). The Effects of Preschool Education What We Know, How Public Policy Is or Is Not Aligned With the Evidence Base, and What We Need to Know. *Psychological Science in the Public Interest*, *10*(2), 49–88.
Pianta, R. C., DeCoster, J., Cabell, S., Burchinal, M., Hamre, B. K., Downer, J., … Howes, C. (2014). Dose–response relations between preschool teachers' exposure to components of professional development and increases in quality of their interactions with children. *Early Childhood Research Quarterly*, *29*(4), 499–508. http://doi.org/10.1016/j.ecresq.2014.06.001
Pinheiro, J. C., & Bates, D. M. (2000). *Mixed effects models in S and S-PLUS*. New York, NY: Springer New York.
Pinheiro, J. C., Bates, D. M., DebRoy, S., Sarkar, D., & R Core Team. (2015). nlme: Linear and Nonlinear Mixed Effects Models. Retrieved from http://cran.r-project.org/package=nlme
PISA Governing Board. (2014). Statement by the PISA Governing Board. Retrieved from https://www.oecd.org/pisa/aboutpisa/statement by the pisa governing board.pdf
Polesel, J. (2010). Vocational education and training (VET) and young people: The pathway of the poor? *Education + Training*, *52*(5), 415–426. http://doi.org/10.1108/00400911011058352
Pons, X. (2012). Going beyond the "PISA Shock" Discourse: An Analysis of the Cognitive Reception of PISA in Six European Countries, 2001-2008. *European Educational Research Journal*, *11*(2).
Popham, W. J. (1988). The dysfunctional marriage of formative and summative teacher evaluation. *Journal of Personnel Evaluation in Education*, *1*(3), 269–273. http://doi.org/10.1007/BF00123822
Popkewitz, T. (2011). PISA: Numbers, Standardizing Conduct, and the Alchemy of School Subjects. In M. A. Pereyra, H.-G. Kotthoff, & R. Cowen (Eds.), *PISA under examination: Changing Knowledge, Changing Tests, and Changing Schools* (pp. 31–46). Rotterdam: SensePublishers.
Popp, M., Knodel, P., Martens, K., & de Olano, D. (2010). Bilanz – Was PISA im Bildungsbereich verändert hat. In P. Knodel, K. Martens, D. de Olano, & M. Popp (Eds.), *Das PISA Echo* (pp. 301–313). Frankfurt am Main: Campus Verlag.
Porter, T. (1995). *Trust in Numbers: the pursuit of objectivity in science and public life*. Chichester: Princeton University Press. http://doi.org/10.1136/bmj.311.7002.460a
Porter, T., & Webb, M. (2007). *The Role of the OECD in the Orchestration of Global Knowledge Networks* (Paper presented at the Global Knowledge Networks, 30 May). Saskatoon, Canada.
Prain, V., Cox, P., Deed, C., Dorman, J., Edwards, D., Farrelly, C., … Yager, Z. (2013). Personalised learning: lessons to be learnt. *British Educational Research Journal*, *39*(4), 654–676. http://doi.org/10.1080/01411926.2012.669747
Prais, S. J. (2003). Cautions on OECD'S Recent Educational Survey (PISA). *Oxford Review*

of Education, *29*(2), 139–163. http://doi.org/10.1080/0305498032000080657
Prais, S. J. (2004). Cautions on OECD's Recent Educational Survey (PISA): Rejoinder to OECD's Response. *Oxford Review of Education*, *30*(4), 569–573. Retrieved from http://www.jstor.org/stable/4127166
Prenzel, M., Kobarg, M., Schöps, K., & Rönnebeck, S. (Eds.). (2013). *Research on PISA*. Dordrecht: Springer Science+Business Media B.V. http://doi.org/10.1007/978-94-007-4458-5
Prokic-Breuer, T., & Dronkers, J. (2012). The high performance of Dutch and Flemish 15-year-old native pupils: explaining country differences in math scores between highly stratified educational systems. *Educational Research and Evaluation*, *18*(8), 749–777. http://doi.org/10.1080/13803611.2012.727359
Quinn, R. (2014). Opening Address. In *Learning from DEIS*. Dublin. Retrieved from http://www.education.ie/en/Schools-Colleges/Services/DEIS-Delivering-Equality-of-Opportunity-%E2%80%8Ein-Schools-/DEIS-Supporting-Information/Learning-from-DEIS-Seminar.html
Quittre, V., Baye, A., Fagnant, A., & Hindryckx, G. (2008). *PISA 2006 - Evaluation de la culture scientifique des jeunes de 15 ans: Document à l'attention des professeurs de sciences des 1er et 2e degrés de l'enseignement secondaire*. Liège: Service d'Analyse des Systèmes et des Pratiques d'Enseignement, Université de Liège. Retrieved from http://www.enseignement.be/download.php?do_id=5061&do_check=
R Core Team. (2015). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org/
Raag, T., Kusiak, K., Tumilty, M., Kelemen, A., Bernheimer, H., & Bond, J. (2011). Reconsidering SES and gender divides in literacy achievement: are the gaps across social class and gender necessary? *Educational Psychology*, *31*(6), 691–705. http://doi.org/10.1080/01443410.2011.599835
Ragin, C. C. (1987). *The comparative method*. London: University of California Press, Ltd.
Raise Your Hand Texas. (2015). School Vouchers: Not Good for Kids, Not Good for Texas. Retrieved from http://www.raiseyourhandtexas.org
Ramey, C. T., & Ramey, S. L. (1998). Early intervention and early experience. *The American Psychologist*, *53*(2), 109–120. http://doi.org/10.1037/0003-066X.53.2.109
Ramey, C. T., Ramey, S. L., & Lanzi, R. G. (2006). Children's health and education. In I. Sigel & A. Renninger (Eds.), *The handbook of child psychology* (pp. 864–892). Hoboken, NJ: John Wiley & Sons, Inc.
Rappleye, J. (2006). Theorizing Educational Transfer: toward a conceptual map of the context of cross-national attraction. *Research in Comparative and International Education*, *1*(3), 223–240. http://doi.org/10.2304/rcie.2006.1.3.223
Rasbash, J., Steele, F., Browne, W. J., & Goldstein, H. (2012). *A User's Guide to MLwiN, v2.26*. Centre for Multilevel Modelling, University of Bristol. Retrieved from http://www.bristol.ac.uk/cmm/software/mlwin/download/manuals.html
Rasch, G. (1980). *Probabilistic Models for Some Intelligence and Attainment Tests* (Expanded). Chicago, IL: University of Chicago Press.
Raudenbush, S. W. (1989). "Centering" predictors in multilevel analysis: Choices and consequences. *Multilevel Modelling Newsletter*, *1*(2), 10–12.
Raudenbush, S. W. (2015). Value Added: A Case Study in the Mismatch Between Education Research and Policy. *Educational Researcher*, *44*(2), 138–141. http://doi.org/10.3102/0013189X15575345
Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models* (2nd ed.). Thousand Oaks, CA: SAGE Publications, Inc.
Rautalin, M., & Alasuutari, P. (2007). The Curse of Success: the impact of the OECD's Programme for International Student Assessment on the discourses of the teaching profession in Finland. *European Educational Research Journal*, *6*(4), 348. http://doi.org/10.2304/eerj.2007.6.4.348
Rautalin, M., & Alasuutari, P. (2009). The uses of the national PISA results by Finnish

officials in central government. *Journal of Education Policy*, *24*(5), 539–556. http://doi.org/10.1080/02680930903131267
Reiss, M. J. (2013). *How can we get more students to study mathematics or physics? – The UPMAP Project* (Research Briefing No. 2). London: Institute of Eudcaion, University College London.
Rey, B., Caffieaux, C., Defrance, A., & Marcoux, G. (2005). L'articulation entre savoirs et compétences dans l'enseignement secondaire. *Informations Pédagogiques*, *57*, 3–14.
Reynolds, D., Teddlie, C., & Creemers, B. (2002). *World class schools*. London: Carfax.
Riehl, C. (1999). Labeling and Letting Go: An Organizational Analysis of How High School Students Are Discharged as Dropouts. In A. M. Pallas (Ed.), *Research in Sociology of Education and Socialization* (pp. 231–268). New York, NY: JAI Press.
Ringarp, J., & Rothland, M. (2010). Is the Grass Always Greener? The Effect of the PISA Results on Education Debates in Sweden and Germany. *European Educational Research Journal*, *9*(3), 422–430.
Rinne, R., Kallo, J., & Hokka, S. (2004). Too Eager to Comply? OECD Education Policies and the Finnish Response. *European Educational Research Journal*, *3*(2), 454–485. http://doi.org/10.2304/eerj.2004.3.2.3
Robertson, S. L., Novelli, M., Dale, R., Tikly, L., Dachi, H., & Alphonce, N. (2007). *Globalisation, Education and Development: Ideas, Actors and Dynamics*. London: DFID Publications.
Robinson, J. P., & Lubienski, S. T. (2011). The Development of Gender Achievement Gaps in Mathematics and Reading During Elementary and Middle School Examining Direct Cognitive Assessments and Teacher Ratings. *American Educational Research Journal*, *48*(2), 268–302. http://doi.org/10.3102/0002831210372249
Robinson-Cimpian, J. P., Lubienski, S. T., Ganley, C. M., & Copur-Gencturk, Y. (2014). Teachers' perceptions of students' mathematics proficiency may exacerbate early gender gaps in achievement. *Developmental Psychology*, *50*(4), 1262–1281. http://doi.org/10.1037/a0035073
Roe, E. (1994). *Narrative Policy Analysis: Theory and practice*. Durham, NC: Duke University Press.
Roemer, J. E. (1998). *Equality of Opportunity*. Cambridge, MA: Harvard University Press.
Rogers, P. J. (2008). Using Programme Theory to Evaluate Complicated and Complex Aspects of Interventions. *Evaluation*, *14*(1), 29–48. http://doi.org/10.1177/1356389007084674
Rogers, P. J., Petrosino, A., Huebner, T. A., & Hacsi, T. A. (2000). Program theory evaluation: Practice, promise, and problems. *New Directions for Evaluation*, *2000*(87), 5–13. http://doi.org/10.1002/ev.1177
Romainville, M. (2002). Du bon usage de PISA. *Revue Nouvelle*, *115*(3-4), 86–99. http://doi.org/10.3917/rmm.091.0033
Romainville, M. (2008). Et si on arrêtait de tirer sur les compétences? *inDirect*, *10*, 31–44.
Romer, P. M. (1990). Endogenous Technological Change. *Journal of Political Economy*, *98*(5), S71–S102. http://doi.org/10.1086/261725
Rose, R., Shevlin, M., Winter, E., & O'Raw, P. (2010). Special and inclusive education in the Republic of Ireland: reviewing the literature from 2000 to 2009. *European Journal of Special Needs Education*, *25*(4), 359–373. http://doi.org/10.1080/08856257.2010.513540
Rossi, P. H., Freeman, H. E., & Lipsey, M. W. (2004). *Evaluation: A systematic approach* (7th ed.). Thousand Oaks, CA: SAGE Publications, Inc.
Rouse, C. E., & Barrow, L. (2009). School Vouchers and Student Achievement: Recent Evidence and Remaining Questions. *Annual Review of Economics*, *1*(1), 17–42. http://doi.org/10.1146/annurev.economics.050708.143354
Rowley, J. (2012). Conducting research interviews. *Management Research Review*, *35*(3/4), 260–271. http://doi.org/10.1108/01409171211210154
Royal Irish Academy Committee for Mathematical Sciences. (2006). NCCA Review of

Mathematics in Post-Primary Education. *Irish Mathematical Society Bulletin*, *57*, 11–20.
Rumberger, R. W. (2003). Consequences of Student Mobility. *Journal of Negro Education*, *72*(1), 6–21.
Rumberger, R. W., & Lim, S. A. (2008). *Why Students Drop Out of School: A Review of 25 Years of Research* (California Dropout Research Project Report No. 15). Santa Barbara, CA: University of California Press.
Rumberger, R. W., & Palardy, G. J. (2005). Test Scores, Dropout Rates, and Transfer Rates as Alternative Indicators of High School Performance. *American Educational Research Journal*, *42*(1), 3–42. http://doi.org/10.3102/00028312042001003
Russell, L. J. (1951). Propositions and proposals. *Australasian Journal of Philosophy*, *29*(3), 155–163. http://doi.org/10.1080/00048405185200181
Rutkowski, D. (2008). Towards an Understanding of Educational Indicators. *Policy Futures in Education*, *6*(4), 470. http://doi.org/10.2304/pfie.2008.6.4.470
Rutkowski, L., & Rutkowski, D. (2009). Trends in TIMSS responses over time: evidence of global forces in education? *Educational Research and Evaluation*, *15*(2), 137–152. http://doi.org/10.1080/13803610902784352
Ryan, S. (1994). *Home-School-Community Liaison scheme: Final Evaluation Report*. Dublin: Educational Research Centre.
Ryelandt, N. (2010). Le Conseil de l'éducation et de la formation. *Courrier Hebdomadaire Du CRISP*, (35). http://doi.org/10.3917/cris.2080.0005
Ryelandt, N. (2013). Les décrets « inscriptions » et « mixité sociale » de la Communauté française. *Courrier Hebdomadaire Du CRISP*, (23-24).
Saar, E., Ure, O. B., & Holford, J. (Eds.). (2013). *Lifelong Learning in Europe: National Patterns and Challenges*. Cheltenham: Edward Elgar.
Sahlgren, G. H. (2015). *Real Finnish Lessons: The true story of an education superpower*. London: Centre for Policy Studies.
Salverda, W. (Ed.). (2011). *Inequalities' Impacts* (GINI State of the Art Review No. 1). GINI Project. Retrieved from http://www.gini-research.org/articles/papers
Scheerens, J., Luyten, H., van den Berg, S. M., & Glas, C. A. W. (2015). Exploration of direct and indirect associations of system-level policy-amenable variables with reading literacy performance. *Educational Research and Evaluation*, *21*(1), 15–39. http://doi.org/10.1080/13803611.2015.1008520
Schervish, M. J. (1996). P Values: What They Are and What They Are Not. *The American Statistician*, *50*(3), 203. http://doi.org/10.2307/2684655
Schillings, P., Hindryckx, G., Dupont, V., Matoul, A., & Lafontaine, D. (2012). *PIRLS 2011: Enquête internationale sur le développement des compétences en lecture des élèves de 4e année primaire* (Summary report). Bruxelles: Service d'Analyse des Systèmes et des Pratiques d'Enseignement, Université de Liège. Retrieved from http://www.enseignement.be/index.php?page=26044&id_fiche=5516&dummy=26588
Schleicher, A. (2007). Can competencies assessed by PISA be considered the fundamental school knowledge 15-year-olds should possess? *Journal of Educational Change*, *8*(4), 349–357. http://doi.org/10.1007/s10833-007-9042-x
Schleicher, A. (2009a). International Benchmarking as a Lever for Policy Reform. In *Change wars* (pp. 97–116). Bloomington, IN: Solution Tree.
Schleicher, A. (2009b). International Comparisons of Student Learning Outcomes. In A. Hargreaves, A. Lieberman, M. Fullan, & D. Hopkins (Eds.), *Second International Handbook of Educational Change* (pp. 485–504). Dordrecht: Springer Netherlands.
Schleicher, A. (2011a). *Building a High-Quality Teaching Profession: Lessons from around the world* (Background Report for the International Summit on the Teaching Profession). Retrieved from http://www2.ed.gov/about/inits/ed/internationaled/background.pdf
Schleicher, A. (2011b, October 14). Chinese lessons. Retrieved December 5, 2013, from http://oecdeducationtoday.blogspot.co.uk/2011/10/chinese-lessons.html

Schleicher, A. (2012, May 23). Hong Kong's success in PISA – One system, many actors. Retrieved December 5, 2013, from http://oecdeducationtoday.blogspot.co.uk/2012/05/hong-kongs-success-in-and-pisa-one.html
Schleicher, A. (2013a). Lessons from PISA outcomes. *OECD Observer, 297*. Retrieved from http://oecdobserver.org/news/fullstory.php/aid/4239/Lessons_from_PISA_outcomes.html
Schleicher, A. (2013b, April 22). Learning from other countries' experiences in education. Retrieved December 5, 2013, from http://oecdeducationtoday.blogspot.co.uk/2013/04/learning-from-other-countries.html
Schleicher, A. (2013c, July 26). Big data and PISA. *Educationtoday*. Retrieved from http://oecdeducationtoday.blogspot.co.uk/2013/07/big-data-and-pisa.html
Schleicher, A. (2013d, August). Attacks on Pisa are entirely unjustified. *TES Magazine*. Retrieved from http://www.tes.co.uk/article.aspx?storycode=6345213
Schleicher, A. (2013e, December 3). What we learn from the PISA 2012 results. *Educationtoday*. Retrieved from http://oecdeducationtoday.blogspot.co.uk/2013/12/what-we-learn-from-pisa-2012-results.html
Schleicher, A. (2013f, December 10). Are the Chinese cheating in PISA or are we cheating ourselves? *Educationtoday*. Retrieved from http://oecdeducationtoday.blogspot.co.uk/2013/12/are-chinese-cheating-in-pisa-or-are-we.html
Schlicht, R., Stadelmann-Steffen, I., & Freitag, M. (2010). Educational Inequality in the EU: The Effectiveness of the National Education Policy. *European Union Politics*, *11*(1), 29–59. http://doi.org/10.1177/1465116509346387
Schlicht-Schmälzle, R., Teltemann, J., & Windzio, M. (2011). *Deregulation of education: What does it mean for efficiency and equality?* (No. 157). *TranState working papers*.
Schlotter, M., Schwerdt, G., & Wößmann, L. (2011). Econometric methods for causal evaluation of education policies and practices: a non-technical guide. *Education Economics*, *19*(2), 109–137. http://doi.org/10.1080/09645292.2010.511821
Schriewer, J., & Martinez, C. (2004). Constructions of internationality in education. In G. Steiner-Khamsi (Ed.), *The Global Politics of Educational Borrowing and Lending* (pp. 29–53). New York, NY: Teachers College Press.
Schweinhart, L. J., & Fulcher-Dawson, R. (2009). Early childhood education. In G. Sykes, B. Schneider, D. N. Plank, & T. G. Ford (Eds.), *Handbook of education policy research* (pp. 876–888). Oxon: Routledge.
Schweinhart, L. J., Montie, J., Xiang, Z., Barnett, W. S., Belfield, C. R., & Nores, M. (2005). *Lifetime Effects: The High/Scope Perry Preschool Study Through Age 40*. Ypsilanti, MI: High/Scope Press.
Schwerdt, G., & Wößmann, L. (2015). *The Information Value of Central School Exams* (IZA discussion paper No. 9122).
Scott, W. R. (1995). *Institutions and organizations*.
Searle, J. R. (1976). A classification of illocutionary acts. *Language in Society*, *5*(1), 1–23. http://doi.org/10.1017/S0047404500006837
Seeber, S., & Lehmann, R. (2013). The Impact of International Large-Scale Assessments on Work-related Educational Monitoring and Policy-making in Germany. *Research in Comparative and International Education*, *8*(3), 342–348. http://doi.org/10.2304/rcie.2013.8.3.342
Sellar, S., & Lingard, B. (2013a). Looking East: Shanghai, PISA 2009 and the reconstitution of reference societies in the global education policy field. *Comparative Education*, *49*(4), 464–485. http://doi.org/10.1080/03050068.2013.770943
Sellar, S., & Lingard, B. (2013b). The OECD and global governance in education. *Journal of Education Policy*, *28*(5), 710–725. http://doi.org/10.1080/02680939.2013.779791
Serder, M., & Jakobsson, A. (2014). "Why bother so incredibly much?": student

perspectives on PISA science assignments. *Cultural Studies of Science Education*, 1–21. http://doi.org/10.1007/s11422-013-9550-3
Service général de l'Inspection. (2009). *Rapport établi par le Service général de l'Inspection au terme de l'année scolaire 2008-2009*. Retrieved from http://www.enseignement.be/index.php?page=24234&navi=103
Service général de l'Inspection. (2010). *Rapport établi par le Service général de l'Inspection au terme de l'année scolaire 2009-2010*. Retrieved from http://www.enseignement.be/index.php?page=24234&navi=103
Service général de l'Inspection. (2012). *Rapport établi par le Service général de l'Inspection au terme de l'année scolaire 2011-2012*. Retrieved from http://www.enseignement.be/index.php?page=24234&navi=103
Shamatov, D., & Sainazarov, K. (2010). *The impact of standardized testing on education quality in Kyrgyzstan: The case of the Program for International Student Assessment (PISA) 2006*. *The Impact of International Achievement Studies on National Education Policymaking*. Elsevier. http://doi.org/10.1108/S1479-3679(2010)13
Shepard, L. A., Flexer, R. J., Hiebert, E. H., Marion, S. F., Mayfield, V., & Weston, T. J. (1995). *Effects of Introducing Classroom Performance Assessments on Student Learning* (CSE Technical Report No. 394). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
Shiel, G., Cosgrove, J., Sofroniou, N., & Kelly, A. (2001a). *Ready for Life? : the literacy achievements of Irish 15-year olds with comparative international data* (Summary Report). Dublin: Educational Research Centre.
Shiel, G., Cosgrove, J., Sofroniou, N., & Kelly, A. (2001b). *Ready for Life? : the literacy achievements of Irish 15-year olds with comparative international data*. Dublin: Educational Research Centre.
Shiel, G., & Kelly, D. (2001). *The 1999 National Assessment of Mathematics Achievement* (Summary Report). Dublin: Educational Research Centre.
Shiel, G., Moran, G., Cosgrove, J., & Perkins, R. (2010). *A Summary of the Performance of Students in Ireland on the PISA 2009 Test of Mathematical Literacy and a Comparison with Performance in 2003*. Dublin: Educational Research Centre.
Shiel, G., & Murphy, R. (2000). *Drumcondra English Profiles*. Dublin: Educational Research Centre.
Shiel, G., Perkins, R., Close, S., & Oldham, E. (2007). *PISA Mathematics: A Teacher's Guide*. Dublin: Department of Education and Science.
Shiel, G., Surgenor, P., Close, S., & Millar, D. (2006). *The 2004 national assessment of mathematical achievement*. Dublin: Educational Research Centre.
Shober, A. F. (2010). *Splintered Accountability*. (A. l. Schneider & H. M. Ingram, Eds.). Albany, NY: State University of New York.
Shumway, R. H., & Stoffer, D. S. (2011). *Time Series Analysis and Its Applications* (3rd ed.). New York, NY: Springer Science+Business Media LLC. http://doi.org/10.1007/978-1-4419-7865-3
Simola, H., & Rinne, R. (2011). Education politics and contingency: Belief, status and trust behind the Finnish PISA miracle. In M. A. Pereyra, H.-G. Kotthoff, & R. Cowen (Eds.), *PISA under examination: Changing knowledge, changing tests, and changing schools* (pp. 225–244). Rotterdam: SensePublishers.
Smith, T. M., & Baker, D. P. (2001). Worldwide Growth and Institutionalization of Statistical Indicators for Education Policy-Making. *Peabody Journal of Education*, *76*(3), 141–152. http://doi.org/10.1207/S15327930PJE763&4_9
Smyth, E., Darmody, M., McGinnity, F., & Byrne, D. (2009). *Adapting to Diversity: Irish Schools and Newcomer Students* (Research Series No. 8). Dublin: ESRI.
Smyth, E., & McCoy, S. (2009). *Investing in education: combating educational disadvantage* (Research Series No. 6). Dublin: ESRI.
Snyder, S. (2013). *The Simple, the Complicated, and the Complex: Educational Reform Through the Lens of Complexity Theory* (No. 96). *OECD Education Working Papers*.

OECD Publishing.
Spelke, E. S. (2005). Sex differences in intrinsic aptitude for mathematics and science? A critical review. *American Psychologist*, *60*(9), 950–958.
Spinath, B., Eckert, C., & Steinmayr, R. (2014). Gender differences in school success: what are the roles of students' intelligence, personality and motivation? *Educational Research*, *56*(2), 230–243. http://doi.org/10.1080/00131881.2014.898917
Spring, J. H. (1976). *The sorting machine: national educational policy since 1945*. New York, NY: Longman.
Springer, M. G., & Balch, R. (2009). Evaluating and Rewarding the Quality of Teachers: International Practices. In S. Sclafani (Ed.), *Evaluating and Rewarding the Quality of Teachers: International Practices* (pp. 65–101). Paris: OECD Publishing. http://doi.org/10.1787/9789264034358-en
Sriprakash, A., & Mukhopadhyay, R. (2015). Reflexivity and the politics of knowledge: researchers as "brokers" and "translators" of educational development. *Comparative Education*, *51*(2), 231–246. http://doi.org/10.1080/03050068.2014.996027
Stack, E. (2005). Whole School Evaluation and Inspiring Professionalism. In *IPPN Conference* (pp. 1–16). Saggart. Retrieved from http://www.ippn.ie/index.php?option=com_mtree&task=viewlink&link_id=2577&cf_id=24
Stake, R. E. (1967). The Countenance of Educational Evaluation. *Teachers College Record*, *68*, 523–540.
Steiner-Khamsi, G. (2003). The Politics of League Tables. *Journal of Social Science Education*, *2*(1), 1–6.
Steiner-Khamsi, G. (2004a). Blazing a trail for policy theory and practice. In G. Steiner-Khamsi (Ed.), *The Global Politics of Educational Borrowing and Lending* (pp. 201–220). New York, NY: Teachers College Press.
Steiner-Khamsi, G. (2004b). Globalization in education: real or imagined? In G. Steiner-Khamsi (Ed.), *The Global Politics of Educational Borrowing and Lending* (pp. 1–6). New York, NY: Teachers College Press.
Steiner-Khamsi, G. (2006). The economics of policy borrowing and lending: a study of late adopters. *Oxford Review of Education*, *32*(5), 665–678. http://doi.org/10.1080/03054980600976353
Steiner-Khamsi, G. (2010). The Politics and Economics of Comparison. *Comparative Education Review*, *54*(3), 323–342. http://doi.org/10.1086/653047
Steiner-Khamsi, G., & Quist, H. O. (2000). The Politics of Educational Borrowing: Reopening the Case of Achimota in British Ghana. *Comparative Education Review*, *44*(3), 272–299. http://doi.org/10.1086/447615
Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R., & Befani, B. (2012). *Broadening the Range of Designs and Methods for Impact Evaluations* (Working Paper No. 38). Department for International Development. Retrieved from http://www.dfid.gov.uk/Documents/publications1/design-method-impact-eval.pdf
Stoet, G., & Geary, D. C. (2013). Sex Differences in Mathematics and Reading Achievement Are Inversely Related: Within- and Across-Nation Assessment of 10 Years of PISA Data. *PLoS ONE*, *8*(3). http://doi.org/10.1371/journal.pone.0057988
Sugrue, C. (2006). A Critical Appraisal of the Impact of International Agencies on Educational Reforms and Teachers' Lives and Work: the case of Ireland? *European Educational Research Journal*, *5*(3), 181. http://doi.org/10.2304/eerj.2006.5.3.181
Sugrue, C. (2008). *The future of educational change: international perspectives*. Oxon: Routledge.
Supplementary Protocol No. 1 to the Convention for European Economic Co-operation on the Legal Capacity, Privileges and Immunities of the Organisation. (1948). 16 April 1948, Paris. Retrieved from https://www.oecd.org/general/supplementaryprotocolno1totheconventionforeuropeaneconomicco-operationonthelegalcapacityprivilegesandimmunitiesoftheorganisation.htm

Surgenor, P., & Shiel, G. (2008). Parity of provision? Learning support for English and mathematics in Irish primary schools. *Irish Educational Studies*, *27*(2), 107–119. http://doi.org/10.1080/03323310802021789
Sylla, N., & De Vos, L. (2006). Facteurs déterminant l'application de nouvelles tâches d'enseignement dans le secondaire supérieur. *Revue Des Sciences de L'éducation*, *32*(2), 377–394. http://doi.org/10.7202/014413ar
Sylva, K. (2014). The role of families and pre-school in educational disadvantage. *Oxford Review of Education*, *40*(6), 680–695. http://doi.org/10.1080/03054985.2014.979581
Taguma, M., Kim, M., Wurzburg, G., & Kelly, F. (2009). *OECD Reviews of Migrant Education: Ireland*. Paris: OECD Publishing.
Takayama, K. (2008). The Politics of International League Tables: PISA in Japan's Achievement Crisis Debate. *Comparative Education*, *44*(4), 387–407.
Takayama, K. (2010). Politics of Externalization in Reflexive Times: Reinventing Japanese Education Reform Discourses through "Finnish PISA Success." *Comparative Education Review*, *54*(1), 51–75. http://doi.org/10.1086/644838
Takayama, K. (2013). Untangling the global-distant-local knot: the politics of national academic achievement testing in Japan. *Journal of Education Policy*, *28*(5), 657–675. http://doi.org/10.1080/02680939.2012.758833
Takayama, K., & Apple, M. W. (2008). The cultural politics of borrowing: Japan, Britain, and the narrative of educational crisis. *British Journal of Sociology of Education*, *29*(3), 289–301. http://doi.org/10.1080/01425690801966360
Task Force on the Physical Sciences. (2001). *Report and Recommendations of the Task Force on the Physical Sciences*. Dublin: Department of Education and Science.
Taylor, S., Rizvi, F., Lingard, B., & Henry, M. (1997). *Educational policy and the politics of change*. Oxon: Routledge.
Teltemann, J. (2014). Achievement vs. Equality: What Determines PISA Performance? In K. Martens, P. Knodel, & M. Windzio (Eds.), *Internationalization of Education Policy: A New Constellation of Statehood in Education?* Basingstoke: Palgrave Macmillan.
The World Bank. (2014). *World Bank GNI per capita Operational Guidelines & Analytical Classifications* (Database). Retrieved from http://siteresources.worldbank.org/DATASTATISTICS/Resources/OGHIST.xls
Thilmany, S. (2011). La discrimination positive en question. *Education & Formation*, *e-295*, 118–128.
Thorn, W. (2009). *International Adult Literacy and Basic Skills Surveys in the OECD Region* (OECD Education Working Papers No. 26). Paris: OECD Publishing.
Timperley, H. (2008). *Teacher professional learning and development* (Educational Practice Series) (Vol. 18). International Bureau of Education.
Timperley, H., Wilson, A., Barrar, H., & Fung, I. (2007). *Teacher Professional Learning and Development: Best Evidence Synthesis Iteration [BES]*. Wellington: Ministry of Education.
Tirivayi, N., Maasen van den Brink, H., & Groot, W. (2013). Group incentives for teachers and their effects on student learning: a systematic review of theory and evidence. *School Effectiveness and School Improvement*, *25*(4), 570–601. http://doi.org/10.1080/09243453.2013.857697
Toulmin, S. E. (1958). *The Uses of Argument*. Cambridge: Cambridge University Press.
Toulmin, S. E., Rieke, R., & Janik, A. (1979). *An introduction to reasoning*. New York, NY: Macmillan.
Travers, J. (2010). Learning support policy for mathematics in Irish primary schools: equal access but unequal needs. *Irish Educational Studies*, *29*(1), 71–80. http://doi.org/10.1080/03323310902884342
Tröhler, D. (2011). Concepts, cultures and comparisons: PISA and the double German discontentment. In M. A. Pereyra, H.-G. Kotthoff, & R. Cowen (Eds.), *PISA under examination: Changing knowledge, changing tests, and changing schools* (pp. 245–257). Rotterdam: SensePublishers.

Tucker, M. S. (Ed.). (2011). *Surpassing Shanghai: An agenda for American education built on the world's leading systems*. Cambridge, MA: Harvard University Press.
Turner, V. (1974). *The Ritual Process: structure and anti-structure*. Harmondsworth: Penguin.
Turner, V. (1975). Symbolic Studies. *Annual Review of Anthropology*, *4*, 145–161. http://doi.org/10.1146/annurev.an.04.100175.001045
Tveit, S. (2014). Educational assessment in Norway. *Assessment in Education: Principles, Policy & Practice*, *21*(2), 221–237. http://doi.org/10.1080/0969594X.2013.830079
Tymms, P. (2004). Are standards rising in English primary schools? *British Educational Research Journal*, *30*(4), 477–494. http://doi.org/10.1080/0141192042000237194
UNESCO-IBE. (2011a). *Indonesia* (7th ed.). *World Data on Education*. Retrieved from http://www.ibe.unesco.org/
UNESCO-IBE. (2011b). Malaysia. In *World Data on Education* (7th ed.). Retrieved from http://www.ibe.unesco.org/
UNESCO-IBE. (2015). World Data on Education: Seventh edition 2010/11. Retrieved from http://www.ibe.unesco.org/en/services/online-materials/world-data-on-education/seventh-edition-2010-11.html
UNESCO-UIS. (2012). *The International Standard Classification of Education: ISCED 2011*. *Comparative Social Research*. UNESCO Institute for Statistics.
UNICEF. (2012). *Equity in Learning?* Geneva: UNICEF Regional Office for CEE/CIS.
Usefulness. (2016). In *Merriam-Webster* (Online). Retrieved from http://www.merriam-webster.com/dictionary/usefulness
Usher, A., & Kober, N. (2011). *Keeping informed about school vouchers: A review of major developments and research*. Center on Education Policy. Retrieved from http://www.cep-dc.org/displayDocument.cfm?DocumentID=369
Utility. (2016). In *Merriam-Webster* (Online). Retrieved from http://www.merriam-webster.com/dictionary/utility
Van Damme, J., Bossens, N., Brans, M., & Fobé, E. (2014). *Comparing expectations of and satisfaction with policy advice of policy makers and advisory council members in nine policy domains in Belgium (Flanders)* (Paper presented at IPSA, Montreal, 19-24 July).
Van de Werfhorst, H. G., & Mijs, J. J. B. (2010). Achievement Inequality and the Institutional Structure of Educational Systems: A Comparative Perspective. *Annual Review of Sociology*, *36*(1), 407–428. http://doi.org/10.1146/annurev.soc.012809.102538
Van den Heuvel-Panhuizen, M., & Drijvers, P. (2014). Realistic Mathematics Education. In S. Lerman (Ed.), *Encyclopedia of Mathematics Education* (pp. 521–525). Dordrecht: Springer Netherlands.
van Langen, A., Bosker, R., & Dekkers, H. (2006). Exploring cross-national differences in gender gaps in education. *Educational Research and Evaluation*, *12*(2), 155–177. http://doi.org/10.1080/13803610600587016
Vandenberghe, V. (2004). Iniquité de l'enseignement en Communauté française: bref état des lieux et examen du rôle du mode de groupement des élèves d'un enseignement secondaire massifié. In M. Frenay & C. Maroy (Eds.), *L'école, six ans après le décret «missions»* (pp. 237–251). Louvain: Presses universitaires de Louvain.
Varley, J., Murphy, C., & Veale, Ó. (2008a). *Science in Primary Schools, Phase 1* (Research Report No. 10). Dublin: NCCA.
Varley, J., Murphy, C., & Veale, Ó. (2008b). *Science in Primary Schools, Phase 2* (Research Report No. 11). Dublin: NCCA.
Varley, J., Murphy, C., & Veale, Ó. (2011). At the Crossroads: The Impact of New Irish Science Curricula on First Year Post-Primary Students. *Research in Science Education*, *43*(1), 275–298. http://doi.org/10.1007/s11165-011-9252-3
Vescio, V., Ross, D., & Adams, A. (2008). A review of research on the impact of professional learning communities on teaching practice and student learning. *Teaching and Teacher Education*, *24*(1), 80–91. http://doi.org/10.1016/j.tate.2007.01.004

Vincent-Lancrin, S. (2008). The Reversal of Gender Inequalities in Higher Education: An On-going Trend. In *Higher Education to 2030* (Vol. 1). Paris: OECD Publishing.
Volante, L., & Ben Jaafar, S. (2008). Educational assessment in Canada. *Assessment in Education: Principles, Policy & Practice*. http://doi.org/10.1080/09695940802164226
von Bogdandy, A., & Goldmann, M. (2008). The Exercise of International Public Authority through National Policy Assessment: The OECD's PISA Policy as a Paradigm for a New International Standard Instrument. *International Organizations Law Review, 5*(2), 241–298. http://doi.org/10.1163/157237408X412907
von Bogdandy, A., & Goldmann, M. (2012). Taming and Framing Indicators: A Legal Reconstruction of the OECD's Programme for International Student Assessment (PISA). In *Governance by Indicators* (pp. 52–85). Oxford: Oxford University Press. http://doi.org/10.1093/acprof:oso/9780199658244.003.0003
von Davier, M., Gonzalez, E. J., & Mislevy, R. J. (2009). What are plausible values and why are they useful? In *IERI monograph series: Issues and methodologies in large-scale assessments* (Vol. 2, pp. 9–36). Hamburg and Princeton, NJ: IEA-ETS Research Institute (IERI).
von Fintel, K. (2006). Modality and Language. In D. M. Borchert (Ed.), *Encyclopedia of Philosophy* (2nd ed.). Detroit, MI: Macmillan. Retrieved from http://web.mit.edu/fintel/www/modality.pdf
Vong, T. S.-K., & Wong, M. (2014). Lost in Translation? A Case Study of Macao in Fabricating a European Education Space in Asia. *European Educational Research Journal*, *13*(3), 350–359. http://doi.org/10.2304/eerj.2014.13.3.350
Waldinger, F. (2006). *Does Tracking Affect the Importance of Family Background on Students' 'Test Scores?* Retrieved from http://cep.lse.ac.uk/conference_papers/Cambridge2006/Waldinger.pdf
Waldow, F., Takayama, K., & Sung, Y.-K. (2014). Rethinking the pattern of external policy referencing: media discourses over the "Asian Tigers"' PISA success in Australia, Germany and South Korea. *Comparative Education*, *50*(3), 302–321. http://doi.org/10.1080/03050068.2013.860704
Walton, D. N. (2001). Abductive, Presumtive and Plausible Arguments. *Informal Logic*, *21*(2), 141–169.
Walton, D. N. (2006a). *Fundamentals of critical argumentation*. New York, NY: Cambridge University Press. http://doi.org/10.1017/CBO9780511807039
Walton, D. N. (2006b). How to make and defend a proposal in a deliberation dialogue. *Artificial Intelligence and Law*, *14*(3), 177–239. http://doi.org/10.1007/s10506-006-9025-x
Waslander, S., Pater, C., & van der Weide, M. (2010). *Markets in Education: An Analytical Review of Empirical Research on Market Mechanisms in Education* (No. 52). *OECD Education Working Papers*. Paris: OECD Publishing.
Wedell, M. (2009). *Planning for educational change: Putting people and their contexts first*. London: Continuum International Publishing Group Ltd.
Weir, S. (2003). *The evaluation of Breaking The Cycle: a follow-up of the achievements of 6th class pupils in urban schools in 2003*. Dublin: Educational Research Centre.
Weir, S. (2004). *A Commentary on the Implementation of Giving Children an Even Break*. Dublin: Educational Research Centre.
Weir, S. (2014). *The evaluation of the DEIS Programme in primary and second level schools* (Presentation to the National Educational Psychological Service). Unpublished manuscript.
Weir, S., & Archer, P. (2004). *A review of school-based measures aimed at addressing educational disadvantage in Ireland*. Dublin: Educational Research Centre.
Weir, S., Archer, P., & Millar, D. (2009). *Educational Disadvantage in Primary Schools in Rural Areas* (No. 1). Dublin: Educational Research Centre.
Weir, S., Archer, P., O'Flaherty, A., & Gilleece, L. (2011). *A Report on the First Phase of the Evaluation of DEIS*. Dublin: Educational Research Centre.

Weir, S., & Denner, S. (2013). *The evaluation of the School Support Programme under DEIS: Changes in pupil achievement in urban primary schools between 2007 and 2013* (Bullettin Report). Dublin: Educational Research Centre.
Weir, S., & Errity, D. (2014). Addressing disadvantage: What have we learned from the evaluation of DEIS in urban primary schools ? In *Learning from DEIS*. Dublin. Retrieved from http://www.education.ie/en/Schools-Colleges/Services/DEIS-Delivering-Equality-of-Opportunity-%E2%80%8Ein-Schools-/DEIS-Supporting-Information/Learning-from-DEIS-Seminar.html
Weir, S., & McAvinue, L. (2012). *The Impact of DEIS on Class Size in Primary Schools*. Dublin: Educational Research Centre.
Weir, S., & McAvinue, L. (2013). *The achievements and characteristics of pupils attending rural schools participating in DEIS*. Dublin: Educational Research Centre.
Weiss, C. H. (1997). How Can Theory-Based Evaluation Make Greater Headway? *Evaluation Review*, *21*(4), 501–524. http://doi.org/10.1177/0193841X9702100405
Weiss, C. H. (2000). Which Links in Which Theories Shall We Evaluate? In *New Directions for Evaluation* (pp. 35–45). San Francisco, CA: Jossey-Bass.
What Works Clearinghouse. (2012). *WWC review of the report: Milwaukee Parental Choice Program longitudinal educational growth study fifth year report*. Retrieved from http://whatworks.ed.gov
Wickham, H. (2009). *ggplot2: elegant graphics for data analysis*. New York, NY: Springer New York. Retrieved from http://had.co.nz/ggplot2/book
Wilkinson, R., & Pickett, K. (2009). *The Spirit Level: Why Greater Equality Makes Societies Stronger*. New York, NY: Bloomsbury Press.
Willms, J. D. (2003). *Student engagement at school: A sense of belonging and participation*. Paris: OECD Publishing.
Willms, J. D. (2006). *Learning Divides: Ten Policy Questions About the Performance and Equity of Schools and Schooling Systems*. Montreal: UNESCO-UIS.
Wiseman, A. W. (2010). Introduction: The advantages and disadvantages of national education policymaking informed by international achievement studies. In *The Impact of International Achievement Studies on National Education Policymaking* (pp. xi–xxii). Elsevier. http://doi.org/10.1108/S1479-3679(2010)0000013003
Witte, J. F. (2009). Vouchers. In G. Sykes, B. Schneider, D. N. Plank, & T. G. Ford (Eds.), *Handbook of education policy research* (pp. 491–501). Oxon: Routledge.
Wong, V. C., Cook, T. D., Barnett, W. S., & Jung, K. (2008). An effectiveness-based evaluation of five state pre-kindergarten programs. *Journal of Policy Analysis and Management*, *27*(1), 122–154. http://doi.org/10.1002/pam.20310
Woodward, R. (2009). *The Organisation for Economic Co-operation and Development (OECD)*. London: Routledge.
Wößmann, L. (2005). The effect heterogeneity of central examinations: evidence from TIMSS, TIMSS-Repeat and PISA. *Education Economics*, *13*(2), 143–169. http://doi.org/10.1080/09645290500031165
Wößmann, L., Lüdemann, E., Schütz, G., & West, M. R. (2007). *School Accountability, Autonomy, Choice, and the Level of Student Achievement* (No. 13). Paris: OECD. Retrieved from http://www.oecd-ilibrary.org/content/workingpaper/246402531617
Wu, M. L. (2009). A comparison of PISA and TIMSS 2003 achievement results in mathematics. *PROSPECTS*, *39*(1), 33–46. http://doi.org/10.1007/s11125-009-9109-y
Wu, M. L. (2010). Measurement, Sampling, and Equating Errors in Large-Scale Assessments. *Educational Measurement: Issues and Practice*, *29*(4), 15–27. http://doi.org/10.1111/j.1745-3992.2010.00190.x
Wuttke, J. (2007). Uncertainties and bias in PISA. In S. T. Hopmann, G. Brinek, & M. Retzl (Eds.), *PISA zufolge PISA - PISA According to PISA* (pp. 241–263). Wien: Lit Verlag.
Yeh, S. S. (2010). The Cost Effectiveness of 22 Approaches for Raising Student Achievement. *Journal of Education Finance*, *36*(1), 38–75. http://doi.org/10.1353/jef.0.0029

Yi-Chong, X., & Weller, P. (2008). “To be, but not to be seen”: Exploring the impact of international civil servants. *Public Administration*, *86*(1), 35–51. http://doi.org/10.1111/j.1467-9299.2007.00706.x

Yildirim, H. H., & Berberoğlu, G. (2009). Judgmental and Statistical DIF Analyses of the PISA-2003 Mathematics Literacy Items. *International Journal of Testing*, *9*(2), 108–121. http://doi.org/10.1080/15305050902880736

Yoon, K. S., Duncan, T., Lee, S. W.-Y., Scarloss, B., & Shapley, K. L. (2007). *Reviewing the evidence on how teacher professional development affects student achievement* (Issues & Answers Report, REL 2007 No. 33). National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest. Retrieved from files/852/Class5_REL_PDLitReview.pdf

Zellman, G. L., Ryan, G. W., Karam, R., Constant, L., Salem, H., Gonzalez, G., … Al-Obaidli, K. (2009). *RAND report: Implementation of the K-12 Education Reform in Qatar’s Schools*. Santa Monica, CA: RAND Corporation.

Zumbo, B. D. (2009). Validity as Contextualized and Pragmatic Explanation , and Its Implications for Validation Practice. In R. W. Lissitz (Ed.), *The Concept of Validity: Revisions, new directions and applications* (pp. 65–83). Charlotte, NC: Information Age Publishing, Inc.